

Clinically Informed Intelligent Classification of Ovarian Cancer Cells by Label-Free Holographic Imaging Flow Cytometry

Daniele Pirone, Beatrice Cavina, Daniele Gaetano Sirico, Martina Mugnano, Vittorio Bianco, Lisa Miccio, Anna Myriam Perrone, Anna Maria Porcelli, Giuseppe Gasparre, Ivana Kurelac, Pasquale Memmolo,* and Pietro Ferraro*

Liquid biopsy, intended as the detection of circulating tumor cells (CTCs) in hematic specimens, is an emerging tool for both early cancer detection and estimation of prognosis. Herein, the strength of quantitative phase imaging (QPI) is investigated to achieve effective distinction of ovarian cancer (OC) from other blood cell populations based on label-free morphological biomarkers rather than conventional fluorescent imaging or other molecular parameters. At this purpose, QPI is implemented in high-throughput flow cytometry mode and combined with machine learning (ML), reliable and accurate OC cell phenotyping is achieved by developing *ad-hoc* multi-level ML classification architectures driven by a priori clinical information. It is shown that the latter allows increasing the overall classification accuracy when compared to noninformed ML classification systems. Thanks to its simplicity, the proposed intelligent system is compatible with various clinical applications, particularly in the context of CTC-based liquid biopsy during patient follow-up, when cancer subtype and other clinical information are already known.

survival rate (41–50%).^[1] Screening protocols capable of detecting early disease, while tumor mass is still confined to the ovary, could significantly improve this inauspicious outcome. Although available data are heterogeneous in terms of experimental design and methodologies applied, it has been shown that CTCs are a promising biomarker for predicting OC survival and chemotherapy response during patient follow-up,^[2–5] whose current protocols include transvaginal ultrasound, computerized tomography (CT), and positron-emitting tomography (PET), together with the measurement of CA125 blood levels. However, the latter provides only indirect information on disease progression and imaging is not always capable to detect metastatic lesions until they reach a sufficient size, issues that might be resolved by

1. Introduction

Ovarian cancer (OC) is most often diagnosed when it has already reached a metastatic stage, resulting in low five-year relative

CTC-based liquid biopsy. Nonetheless, the transfer readiness level of using OC CTCs as clinical markers is still relatively low, mainly due to methodological drawbacks related to their detection.^[6]

D. Pirone, D. G. Sirico, M. Mugnano, V. Bianco, L. Miccio, P. Memmolo, P. Ferraro
 Institute of Applied Sciences and Intelligent Systems “Eduardo Caianiello”, CNR-ISASI
 Via Campi Flegrei 34, 80078 Pozzuoli (Napoli), Italy
 E-mail: pasquale.memmolo@isasi.cnr.it; pietero.ferraro@isasi.cnr.it

B. Cavina, A. M. Perrone, G. Gasparre, I. Kurelac
 Department of Medical and Surgical Sciences, DIMEC
 Alma Mater Studiorum-University of Bologna
 40138 Bologna, Italy


A. M. Perrone
 Division of Oncologic Gynecology
 IRCCS Azienda Ospedaliero-Universitaria di Bologna
 40138 Bologna, Italy

A. M. Porcelli
 Department of Pharmacy and Biotechnology (FABIT)
 University of Bologna
 40126 Bologna, Italy

A. M. Porcelli, G. Gasparre, I. Kurelac
 Centre for Applied Biomedical Research (CRBA)
 University of Bologna
 40138 Bologna, Italy

A. M. Porcelli, G. Gasparre, I. Kurelac
 Centro Studi e Ricerca Sulle Neoplasie Ginecologiche (CSR)
 University of Bologna
 40138 Bologna, Italy

A. M. Porcelli
 Interdepartmental Center of Industrial Research (CIRI) Life Science and Health Technologies
 University of Bologna
 40064 Ozzano dell'Emilia, Italy

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/aisy.202400390>.

© 2024 The Author(s). Advanced Intelligent Systems published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/aisy.202400390

Recently, increasing attention has been drawn on label-free techniques for the detection and isolation of CTCs,^[7–13] including intelligent lab-on-a-chip platforms in a microfluidic environment.^[14] Instead of relying on a molecular marker, these technologies distinguish CTCs from blood cells by exploiting their biophysical properties, such as size, deformability, and density.^[15] In this context, microfluidic flow cytometry coupled to quantitative phase imaging (QPI) may be envisioned as a valid label-free method for CTC-based liquid biopsy. Indeed, it would offer simultaneous identification and quantification of cancer cells, as well as the high-throughput feature required by routine clinical protocols. QPI measures the complete morphological information of cell shape, as well as intracellular organelle organization, by detecting an intrinsic endogenous marker represented by the refractive index (RI), thus bypassing canonical molecular analyses or chemical staining processes.^[16–19] However, while the morphology of most peripheral blood cells is substantially different from CTCs, the latter share a similar biophysical signature with monocytes. Correct distinction between these two entities, CTCs and monocytes, is one of the main challenges in the context of oncologic label-free liquid biopsy. A remarkable solution has been demonstrated by combining microscopy imaging with the framework of image-based classification through artificial intelligence (AI).^[20–22] Indeed, concurrent with single-cell technologies, machine learning (ML) has redefined the capability to analyze large-scale data by using sophisticated classification methods. The power of ML has recently been demonstrated in several areas of single-cell analysis field.^[23–26] In the context of liquid biopsy, intelligent decision makers have been developed, which are able to efficiently identify CTCs among the background of white blood cells in a fluorescence-based imaging flow cytometry apparatus.^[27,28] Recently, we have investigated the use of an ML-powered QPI flow cytometry (QPI-FC) technique to classify tumor cells and monocytes^[29,30] and phenotype neuroblastoma.^[31] Both approaches addressed the issue of the identification of CTCs with respect to monocytes in early cancer detection, thus demonstrating the effective potential of QPI-FC as candidate technological solution for label-free liquid biopsy.

Following the path opened by recent works on AI applications in OC,^[32–34] in this article we propose a generalized version of the ML-powered QPI-FC to remarkably increase reliability in the identification of OC CTCs when patient follow-up stage is considered. The core of the method proposed consists in designing a three-level hierarchical structure of the classifier that is informed by a priori available diagnostic data about the OC phenotypes to be identified, in order to improve its classification accuracy. At the first classification level, the OC CTCs identification with respect to the background of white blood cells is performed, as in the scenario of early cancer detection. Here, we introduce additional phenotypic information about the cancer cell subtypes to design multiple alternatives for the second classification level able to distinguish OC cell lines based on, for example, cancer cell origin, histological type or molecular marker signature. In the proposed ML-powered approach, the prediction of the OC phenotypes is performed at the third level of classification. The performance of the presented strategy is compared with two-level hierarchical decision-maker where the second level is created by distinguishing the tumor cell phenotypes without

using a priori information. Such two-level hierarchical decision-maker is considered as the baseline. We focus our investigation on five OC phenotypes for modelling different cancer subtypes and we demonstrate that our phenotypically informed hierarchical decision structuring increases the performance of cancer cell prediction up to 18.3% if compared to the accuracy of the baseline approach. Our results allow to efficiently select and exploit this approach driven by prior clinical information to increase the identification rate of the expected OC CTC phenotype. This outcome could be of particular interest during the patients' follow-up stage, thus opening new perspectives in label-free liquid biopsy.

2. Experimental Section

2.1. Cell Lines

Human OC A2780 cell line was purchased from Sigma Aldrich-MERCK (#93112519) while SKOV3 (#ATCC-HTB-77), OVCAR3 (#ATCC-HTB-161) and CAOV3 (#ATCC-HTB-75) from ATCC. OC314, THP-1 monocyte cell line and Jurkat peripheral blood-derived human T-cell leukemia cell line were supplied by a third part. All cells were authenticated by AmpFISTR Identifier PCR Amplification kit (Applied Biosystems #4322288), using Peak Scanner software 2 (Applied Biosystems) to determine microsatellite alleles. Cellosaurus tool was used for similarity comparison,^[35] computing the score by selecting Tanabe algorithm and “non-empty marker” mode. Our models (Query) shared 85–100% microsatellite similarity with expected cell lines (Figure S1, Supporting Information). Cell line *TP53* mutation status was confirmed by Sanger Sequencing as previously described.^[36] All cells were cultured at 37 °C in an incubator with 5% CO₂. OC cell lines were grown as a monolayer and cultured in RPMI 1640 Medium (Life technologies #31870-025) supplemented with 10% FBS (Life Technologies #10270), 1% Penicillin/Streptomycin (Life Technologies #15070-063) and 2 mM L Glutamine (Lonza BE #17-605E). THP-1 and Jurkat cell lines were cultured in suspension in vertically positioned tissue-culture flasks (Corning, #353018), grown in RPMI 1640 Medium (Life Technologies, #31870-025), supplemented with 10% FBS (Life Technologies #10270), 2 mM L-Glutamine (Lonza, Cat #BE17-605E) and 1% Penicillin/Streptomycin (Lonza, Cat #DE17-602E), and maintained at 37 °C in a humidified atmosphere with 5% CO₂. All cells were regularly tested for mycoplasma contamination upon thawing and prior to the experiments. EVOS M5000 Imaging System (ThermoFisher Scientific #AMF5000) was used to obtain 2D images of the OC cells.

To obtain cell suspensions for the experiments, THP-1 and Jurkat cells were harvested from the cell culture flask, centrifuged for 5 min at 1500 rpm and resuspended in PBS solution containing 10% FBS (10 mL). The OC cells were detached following standard protocols. In detail, the cells were seeded in a flask (Corning, #353018) to reach 70–90% confluency. On the day of the experiment the medium was removed, the cells were washed twice with PBS (10 mL for each wash) and then incubated for 5 min at 37 °C with Trypsin-EDTA solution (2 mL, Sigma, #T4049). Subsequently, PBS containing 10% FBS (8 mL) was

added to block trypsin activity. For all cell lines, prior to the experiment the viability was assessed using Trypan Blue dye (Sigma #T8154) following manufacturer's instructions. In detail, the cell suspension (0.05 mL) was diluted 1:2 with Trypan Blue dye and 0.01 mL was loaded in the Burkner counting chamber to count dead (blue) and live (transparent) cells. Only live cells were considered for experimental cell suspension preparations of 4×10^5 cells mL⁻¹ in PBS solution containing 10% FBS and 300 μ L were injected into the microfluidic channel.

2.2. QPI-FC Recording System and Numerical Processing

A digital holography (DH) microscope based on an off-axis Mach-Zehnder optical interferometer was employed for imaging experiments, as sketched in Figure 1A.^[37] A polarizing beam splitter (PBS) divides the light beam generated by a solid-state continuous wave laser source (Laser Quantum Torus 532 - $\lambda = 532$ nm) into a reference and an object beam. Two half-wave plates (HWP) located in front of and behind the PBS adjust the splitting ratio of the two beams. The object beam, after illuminating the biological sample, is collected by a microscope objective (MO₁) and sent to a tube lens (TL₁). Then, it interferes with the reference beam (in turn collected by a MO₂ and a TL₂) within a beam splitter cube (BS). The resulting interference pattern propagates up to the CMOS recording camera (Genie Nano-CXP Camera, 5120 \times 5120 pixels, $\Delta x = \Delta y = 4.5$ μ m pixel size), which records the digital hologram (see Figure 1B). The recorded field of view (FOV) measures 640 \times 640 μ m². The biological sample is made of single cells flowing and rotating in suspension

along a commercial microfluidic channel (Microfluidic - ChipShop 10000107, 200 μ m \times 1000 μ m cross-section) thanks to the laminar flow produced by a low-pressure pump module (Cetoni NEMESYS 290 N) pushing at 75 nL s⁻¹. According to the reference system sketched in Figure 1B, cells flow along to the y -axis, rotate around the x -axis, and are recorded along the optical z -axis.

For each flowing and rotating cell, hundreds of holograms are recorded. Within each hologram, a region of interest (ROI) was selected around each cell (see Figure 1B), in order to reconstruct the corresponding QPM.^[37] At this aim, the holographic ROI was demodulated by exploiting the DH off-axis configuration, which allows filtering the real diffraction order in the Fourier spectrum.^[38] As cells are recorded out-of-focus in any position along the optical z -axis, a numerical refocusing method (based on the Tamura Coefficient) was applied to compute their in-focus distances.^[39] Then, the angular spectrum formula was implemented to compute the in-focus complex field.^[38] Residual aberrations were removed from its argument by subtracting a reference hologram, and the resulting phase-contrast map was denoised and unwrapped,^[37] thus obtaining the final QPM. In Figure 1C, a QPM for each of the cell lines herein analyzed is shown as example.

2.3. Feature Extraction

For each QPM, 44 features were measured. Among them, 11 features were related to the cell phase values, 9 features were related to the cell morphology, and 24 features were related to the

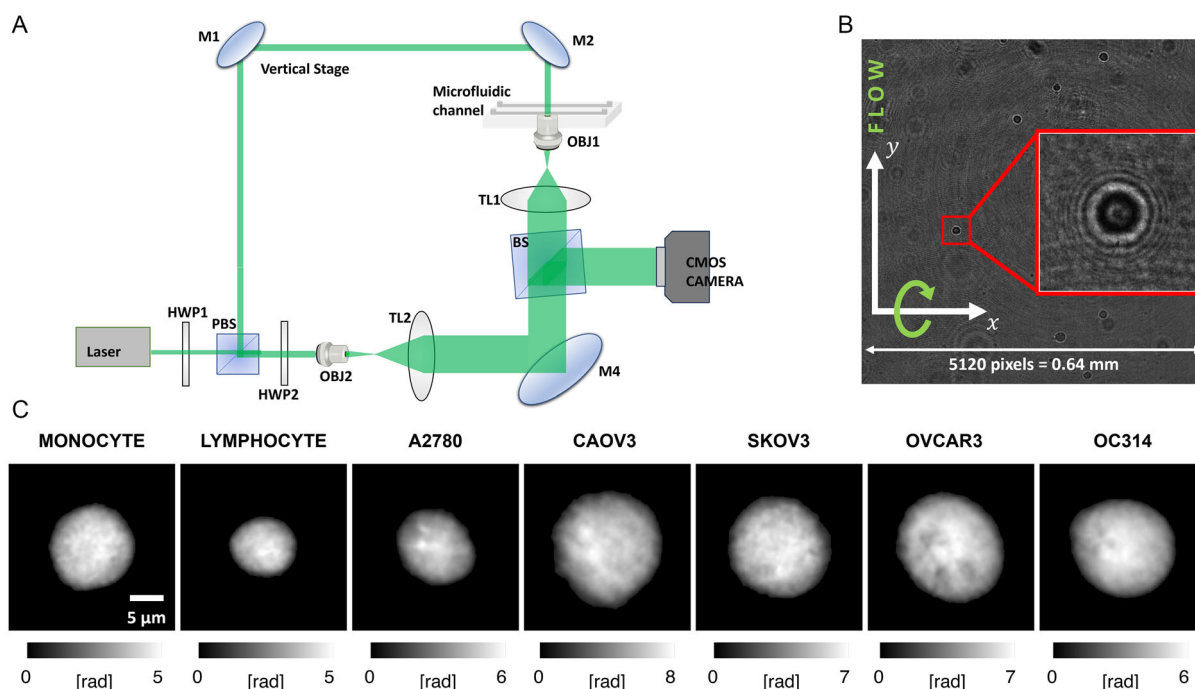


Figure 1. QPI-FC imaging based on DH microscopy. A) Sketch of the opto-fluidic recording system. HWP – Half-wave plate; PBS – Polarizing beam splitter; L1, L2 – Lens; M – Mirror; MO – Microscope objective; MC – Microfluidic channel; TL – Tube lens; BS – Beam splitter; CMOS – Camera. B) Example of a recorded digital hologram, with highlighted in red a holographic ROI, in which the cell's diffraction is visible. C) Example of QPMs for each analyzed cell line.

Gray-Level Co-Occurrence Matrix (GLCM).^[40] The 11 phase features were the average value, the standard deviation, the median, the mode, the maximum value, the skewness, the entropy, the kurtosis, the 0.25 quantile and the 0.75 quantile of the QPM, and the dry mass. In particular, dry mass is the amount of non-aqueous material contained inside the cell and is considered one of the most relevant features accessed by QPI.^[16] The 9 morphological features were the cell area, extent, solidity, circularity, eccentricity, maximum and minimum Feret diameters, major axis length, and normalized centroids distance, computed through the *regionprops* MATLAB function. In particular, given A and P as the area and the perimeter of the cell, respectively, circularity was computed as $4A/P^2$, with 1 being the value of a perfect circle. Having considered an ellipse with the same normalized second central moments as the cell, the major axis length was considered, whereas the eccentricity was the ratio of the distance between the ellipse's foci and its major axis length. The maximum and minimum distances between any two boundary points on the antipodal vertices of the cell convex hull corresponded to the maximum and minimum Feret diameters. Finally, the normalized centroids distance was the distance between the centroid and the weighted centroid (computed with respect to the phase spatial distribution) of the cell, normalized to the radius of a circle having the same area as the cell (i.e., the equivalent radius). With respect to the GLCM, it described how the grey levels are combined among them within an image. In particular, the GLCM $G(i, j, \theta, d)$ measured the number of times a i -pixel occurs at distance d to a j -pixel along the direction θ . We computed 12 GLCM features by averaging the 12 Haralick features (i.e., energy, entropy, correlation, contrast, variance, sum average, inertia, cluster shade, cluster tendency, homogeneity, maximum probability, and inverse variance) at the 0° , 45° , 90° , and 135° directions by fixing a distance $d = 0.5 \mu\text{m}$, while the other 12 GLCM features were computed in the same way but by fixing a distance $2d = 1 \mu\text{m}$.^[41]

3. Results

3.1. QPI-FC ML-Aided Analysis Efficiently Discriminates Various OC Subtypes from Monocytes and Lymphocytes

We recently showed how ML-aided analysis of QPI-FC images is able to discriminate A2780 OC cells from the THP1 monocytic cell line.^[42] Here we expanded our analysis to additional OC models, namely CAO3, SKOV3, OVCAR3, and OC314, with the aim to demonstrate that our approach is capable of correctly classifying OC cells with different phenotypes and/or genotypes from monocytes. Of note, CAO3, OVCAR3, and OC314 recapitulated the high grade serous ovarian cancer (HGSOC), the most common and deadly OC subtype. Moreover, apart from analyzing THP1 cells, we obtained data from the lymphocytic Jurkat cell line, in order to prove that QPI-FC correctly distinguishes cancer cells from lymphocytes, since oncologic patients present with high abundance of this bloodstream cell population. Hence, in this classification problem, seven cell lines were considered, that is, monocytes (MCs) and lymphocytes (LCs) as white blood cells (WBCs), and A2780, CAO3, SKOV3, OVCAR3, and OC314 as ovarian cancer cells (OCCs).

The number of cells recorded through the QPI-FC system is summarized in Table S1 (Supporting Information). The entire dataset was split into an overall training set and a test set with a ratio of 3:1 per cell line (Table S1, Supporting Information).

To solve the single-cell classification issue based on the identification of one out of the seven possible cell lines here considered, four ML hierarchical classifiers (HCs) were built and tested (Figure 2), as discussed below. As sketched in Figure 2, all the HCs share level 1, that is, the classification between WBCs and OCCs. For training each single classifier composing the four HCs, the training set was further extracted from the overall training set summarized in Table S1 (Supporting Information) in order to balance the analyzed classes, as reported in Table S2–S5 (Supporting Information). Moreover, thanks to the cellular roto-translation along the microfluidic channel, hundreds of quantitative phase maps (QPMs) per cell were recorded at different viewing angles (see Experimental Section). This constitutes a relevant difference with respect to conventional QPI-FC systems, in which one image per cell is commonly acquired. A 2D QPM can be considered as the integral of the 3D volumetric distribution of the cell's RIs along the optical axis.^[16] Therefore, two QPMs of the same cell recorded along two different directions are slightly different from each other. Hence, the numerical data augmentation, that is commonly employed to increase the dataset dimension, was hereby replaced by the experimental recording of several QPMs per cell from different points of view. For this reason, over several trainings, 100 QPMs were randomly chosen for each cell of the training set reported in Table S2–S5 (Supporting Information), thus resulting in a $100\times$ data augmentation. Finally, for each QPM, 44 features were measured, that is, 11 phase features 9 morphology features, and 24 GLCM features (see Experimental Section).

To evaluate the performance of level 1 in a more reliable way, it was trained and tested 10 times using the 44 features described. In particular, every time the overall dataset of single cells was randomly divided into an overall training and a test set according to Table S1 (Supporting Information). Subsequently, the several training sets were randomly drawn from the overall training set according to Table S2–S5 (Supporting Information). It is important to underline that, while 100 QPMs per cell were considered for the training sets, all the cell QPMs were instead considered for the test sets. As summarized in Table S6 (Supporting Information), the best ML model was identified for the classification task at level 1, that is, the logistic regression,^[43] by evaluating the classification performance over the test sets through the MATLAB Classification Learner. Hence, the classification performances herein reported must be considered as the average of all 10 evaluations over the corresponding 10 test sets. Performances were quantified by means of three different parameters, that is, recall (REC), precision (PREC), and accuracy (ACC), that are defined as

$$\begin{aligned} \text{REC} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{PREC} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\ \text{ACC} &= \frac{\text{TP}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \end{aligned} \quad (1)$$

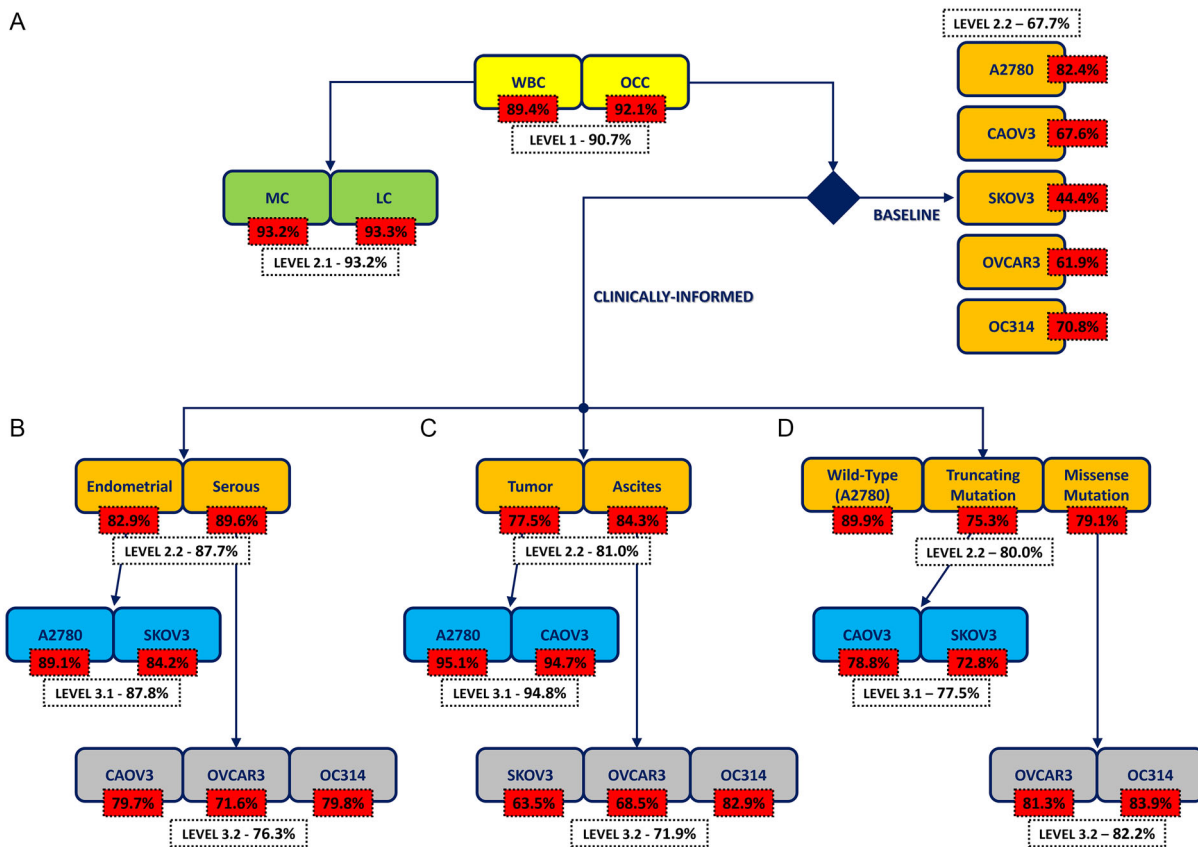


Figure 2. Sketch of the HCs with the average recalls (red boxes) and the average accuracies (white boxes) of each ML classifier, considered independently of each other. A) Non-informed HC. B) Histology-informed HC. C) Origin-informed HC. D) Mutation-informed HC.

where TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively. The accuracy of a classifier is defined as the likelihood of correctly classifying input elements. It is calculated as the fraction of correctly classified items out of the total items classified. Recall for a specific class refers to the probability that its elements are accurately identified by the classifier. This is calculated as the ratio of true positives (correctly identified elements of the class) to the sum of true positives and false negatives (class elements incorrectly not identified). Precision for a particular class, on the contrary, is the proportion of elements that are correctly classified among those categorized as belonging to that class. It is determined by the ratio of true positives to the sum of true positives and false positives (elements not belonging to the class but incorrectly identified as such). Hence, as stated in Equation (1), high recall values indicate a low number of False Negatives (FN), while high precision values suggest a low number of False Positives (FP). Typically, in problems involving cancer diagnosis, the goal is to minimize FN, thereby maximizing the recall value. This approach prioritizes ensuring that as many actual cases of cancer as possible are correctly identified, even if it means accepting a higher number of FP.

Therefore, we report in Figure 2 the average recall (red box) and the average accuracy (white box) related to level 1, that are the same for each of the four HCs. In particular, the classification at

level 1 between WBCs and OCCs has an average accuracy of 90.7%, with the WBCs having an average recall of 89.4% and the OCCs having an average recall of 92.1%.

In summary, our data show that OC and peripheral blood cell processing with QPI-FC and the subsequent AI-mediated analysis of the obtained images successfully distinguish various OC types from MCs and LCs. In the clinical setting, among various tumor-specific features, complete blood count is an important parameter taken into consideration when managing oncologic patients. Thus, we evaluated the ability of our QPI-FC system to correctly distinguish MCs from LCs. As sketched in Figure 2, after identifying WBCs at the level 1, the level 2.1 is accessed, which aims at distinguishing MCs from LCs, and is also common to all the four HCs. The quadratic support vector machine^[43] proved to give the best classification performance with regard to level 2.1 (see Table S6, Supporting Information). In particular, this model was firstly trained and tested 10 times by randomly extracting, every time, the overall training set and the test set from the overall dataset according to Table S1 (Supporting Information), and then by randomly extracting the training set from the overall training set according to Table S2–S5 (Supporting Information). As reported in Figure 2, MCs and LCs are distinguished with a high average accuracy (93.2%), average recall being 93.2% and 93.3%, respectively.

3.2. Identification of Specific OC Subtypes via QPI-FC ML-Aided Analysis Is Optimized by Application of a priori Priors/clinical information

After detecting an OCC at the level 1 of the HC, a more in-depth analysis about the OC phenotype can be performed in order to identify the specific cell line, that is, A2780, CAOV3, SKOV3, OVCAR3, or OC314, which are modelling various clinical subtypes of OC. To do this, there are two possibilities. The first classification case is implemented when there is no a priori information about the type of CTC detected within the blood sample, for example in the scenario of using liquid biopsy as a non-invasive tool for the early diagnosis of cancer. Hence, in this case, the non-informed HC sketched in Figure 2A is employed, in which the level 2.2 is made of a 5-class classifier, avoiding any intermediate clustering of the OCCs. As summarized in Table S6 (Supporting Information), the best ML model at the level 2.2 of the non-informed HC is the linear support vector machine,^[43] which, after being trained and tested 10 times (see Table S1 and S2, Supporting Information), provided an average accuracy of 67.7%, as reported in Figure 2A along with the average recalls of the five OC cell lines.

The second classification case is instead implemented when there is some a priori information about the type of CTC we are searching in the blood specimen, such as in the clinical context of oncologic patient follow-up, when the initial diagnosis and molecular/histology data are most often already available. Thus, we next tested whether such a priori information may be exploited to improve ML-aided identification of CTCs in the patient bloodstream with respect to the non-informed HC. To simulate the clinical data usually available upon diagnosis, we grouped the OCCs in three different subtype categories (see Table S7, Supporting Information), based on the following information available from the literature: 1) histology, 2) origin, and 3) their *TP53* genotype, since mutations in this tumor suppressor gene are the main driver events in ovarian carcinogenesis.^[35,44–46] 1) In the histology category, we distinguished HGSOC (i.e., OVCAR3, CAOV3, and OC314) *versus* endometrial/clear cell OC (E/CCOC)-derived cell lines (i.e., SKOV3 and A2780). Of note, the older literature considered SKOV3 and A2780 as HGSOC, but more recent molecular and histological analyses suggested these cell lines most likely derive from E/CCOC.^[44–46] Indeed, when comparing their phenotype in 2D culture, OVCAR3, CAOV3 and OC314 tend to present with large, polygonal and flattened epithelial morphology, whereas A2780 and SKOV3 more often show a spindle-like shape (Figure S2, Supporting Information). 2) Within the origin category, we differentiated primary tumor-derived (A2780 and CAOV3) *versus* ascites-derived cell lines (SKOV3, OVCAR3, and OC314), the latter being indicative of advanced metastatic disease. 3) The category based on *TP53* mutational status distinguished wild-type (A2780), truncating mutation (SKOV3 and CAOV3) or missense mutation (OVCAR3 and OC314)-bearing cell lines (Table S7, Supporting Information).

Therefore, in addition to the non-informed HC (Figure 2A), three additional ML-based HCs were built and tested, that are here defined as clinically informed HCs, as each of them exploits one of the three a priori phenotypical information about the OC

cells described above. The non-informed HC and the three clinically informed HCs share the level 1, that is, the classification between WBCs and OCCs, and the level 2.1, that is, the classification between MCs and LCs (Figure 2). Instead, the OCC branch changes among these four HCs. With regard to the clinically informed HCs: 1) The histology-informed HC sketched in Figure 2B is based on the preliminary clustering at level 2.2 of OCCs into two groups according to their putative histology, that are E/CCOC cell lines (i.e., A2780 and SKOV3) *versus* HGSOC cell lines (i.e., CAOV3, OVCAR3 and OC314). Next, in order to identify the specific OC cell line, the detection of an endometrial cell opens to the level 3.1, while the detection of a serous cell opens to the level 3.2. 2) The origin-informed HC sketched in Figure 2C is based on the preliminary clustering at level 2.2 of OCCs into two groups according to their origin, i.e., primary tumor-derived cell lines (i.e., A2780 and CAOV3) *versus* ascites-derived cell lines (i.e., SKOV3, OVCAR3 and OC314). Next, in order to identify the specific OC cell line, the detection of tumor-derived cell opens to the level 3.1, while the detection of an ascites-derived cell opens to the level 3.2. 3) The mutation-informed HC sketched in Figure 2D is based on the preliminary clustering of OCCs into three groups according to the *TP53* mutation status, i.e., wild-type cells (A2780) *versus* cells with truncating mutation (CAOV3 and SKOV3) *versus* cells with missense *TP53* mutations (OVCAR3 and OC314). Next, in order to identify the specific OC cell line, the detection of a truncating mutation bearing cell opens to the level 3.1, while the detection of a cell carrying missense mutation opens to the level 3.2.

The single classifiers belonging to the OCC branches of each clinically informed HC architecture were trained through the 44 features computed from the 100 QPMs related to each cell belonging to 10 training sets, extracted according to Table S3–S5 (Supporting Information) from the overall training sets reported in Table S1 (Supporting Information). The OCC classification performances were computed over all the QPMs belonging to the 10 test sets extracted according to Table S1 (Supporting Information). The best ML model was identified again through the MATLAB Classification Learner for each OC classification task inside the three clinically informed HCs, including linear discriminant analysis, logistic regression, support vector machine, and k-nearest neighbors.^[43] A list containing the selected ML models for each classification task is reported in Table S6 (Supporting Information), along with the corresponding model hyperparameters and the possible pre-processing of the dataset based on the z-score standardization (meaning that each feature vector is shifted to have mean 0 and scaled to have standard deviation 1). Therefore, we reported in Figure 2B–D the average recall (red box) and the average accuracy (white box) for each OC classifier, considered independently from the others.

In a HC model, what occurs at a certain classification level depends on what happened at the previous levels, as they are concatenated. For example, considering the OCC branch of the HCs presented in Figure 2, at the level 1, an OCC is correctly recognized with 92.1% accuracy; only the OCCs that were correctly classified at level 1 can access level 2.2, which in turn correctly recognizes endometrial and serous OCCs with an 87.7% accuracy; finally, only the endometrial OCCs that were correctly classified at level 2.2 can access level 3.1, which in turn correctly recognizes A2780 OCCs with a 89.1% probability. However, the

real A2780 recall must also consider the misclassification errors made at level 1 and level 2.2. Thus, the A2780 recall within the histology-informed HC drops to 66.8%. Hence, by considering the several connections within the four HCs, we computed the corresponding confusion matrices displayed in **Figure 3**, averaged among the 10 test sets. This is possible because, within each of the 10 HC trainings, the test set was made of exactly the same cells throughout the several levels of the HC, according to the numbers listed in Table S1 (Supporting Information).

While performance about the classification of MCs and LCs are the same among the several HCs, by comparing the recall values about the OC cell lines reported in Figure 3, it is evident that the employment of an a priori clinical information

(Figure 3B–D) allows improving the performance of the HCs with respect to the baseline, that is, the non-informed HC (Figure 3A). For example, the best recall value in the non-informed HC is reached by CAOV3 cells (i.e., 67.3%), which increases up to 72.5% in the histology-informed HC and 67.9% in the origin-informed HC. Instead, in both the origin-informed HC and mutation-informed HC, the best recall value is reached by the A2780 cells (i.e., 68.6% and 72.7%, respectively), which are higher than the corresponding 66.7% recall of the non-informed HC. However, in the origin-informed HC, the best F1 score (that is the harmonic mean between recall and precision) is obtained with the CAOV3 cells (i.e., 70.8%). More in general, by comparing the four HCs, the best accuracy is related to the histology-informed HC, i.e. 75.1%.

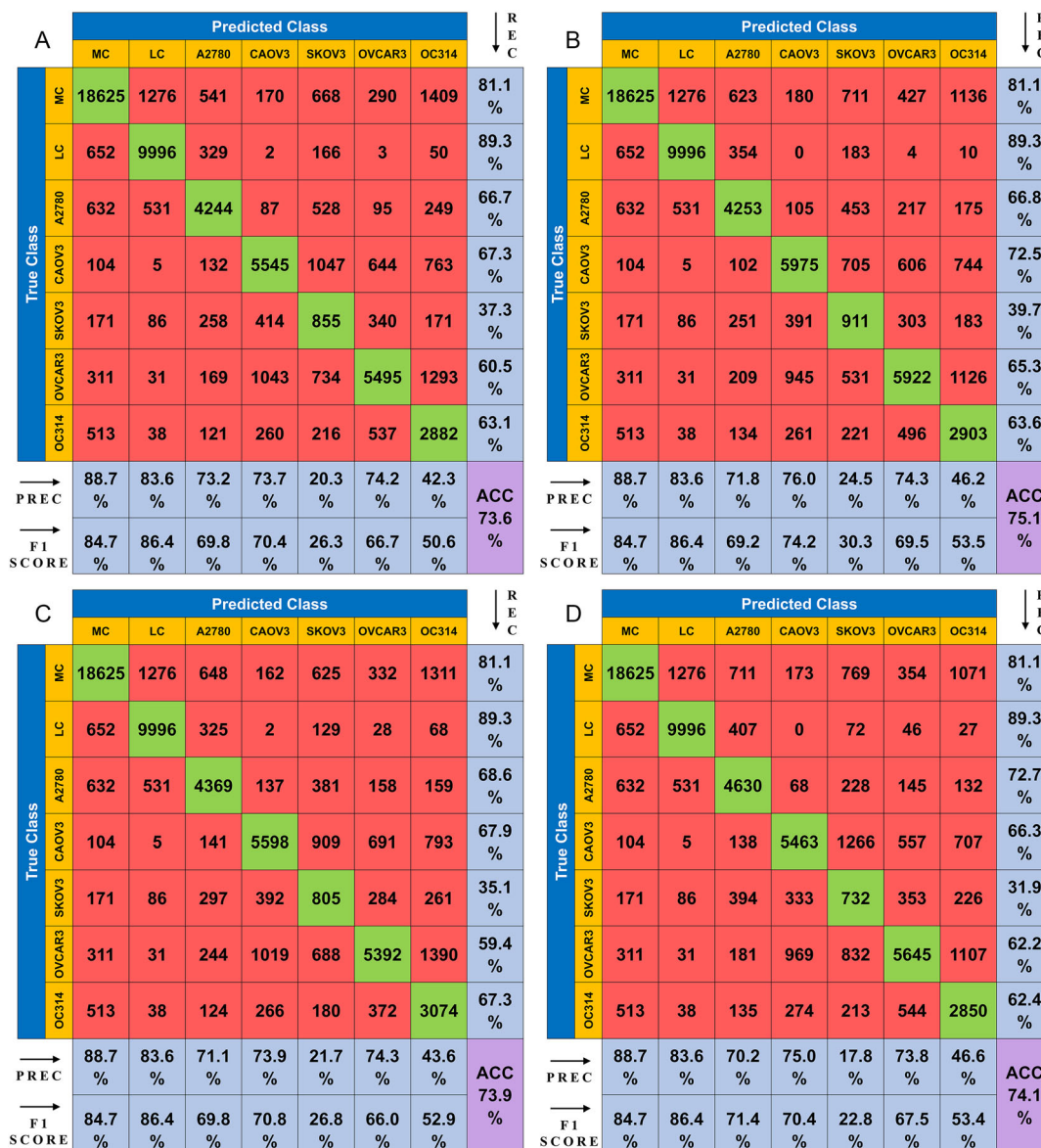


Figure 3. HC confusion matrices related to the QPM classification (before max-voting). Details about the several ML model types used in the four HCs are reported in Table S6 (Supporting Information). A) Non-informed HC. B) Histology-informed HC. C) Origin-informed HC. D) Mutation-informed HC. The bottom row contains the precision values. The right column contains the recall values. The bottom right value (violet) is the HC overall accuracy.

3.3. A Max-Voting Strategy Based on the Cell Rotation Increases the Performance of the Hierarchical Classifiers

The recall values summarized in Figure 3 can be further increased by exploiting the rotation of the cells along the microfluidic channel during the QPI-FC experiments. So far, thanks to the cell rotation, the several QPMs belonging to the same cell were considered as independent elements, i.e., as if each QPM were a different cell, in order to augment the dataset. However, within the QPM dataset, hundreds of QPMs are related to the same cell, even if they were recorded along different viewing angles. Therefore, in addition to the data augmentation for the training step, this property can be further exploited during

the inference step. Indeed, after training a specific HC model, hundreds of QPMs per cell may be recorded thanks to the employed QPI-FC system, which means that hundreds of inferences may be performed per cell, with each inference providing a certain output class. To assign a unique label to the cell, a max-voting strategy can be employed,^[26] i.e., the cell can be classified as belonging to the class that occurred most times across all QPMs inferences.^[31] In this way, we obtained the HC confusion matrices related to the cell classification reported in **Figure 4** for each of the four HCs we assessed. By comparing the QPM confusion matrices in Figure 3 with the corresponding items in Figure 4, an overall increase of the classification performances can be inferred, due to the max-voting strategy that allows

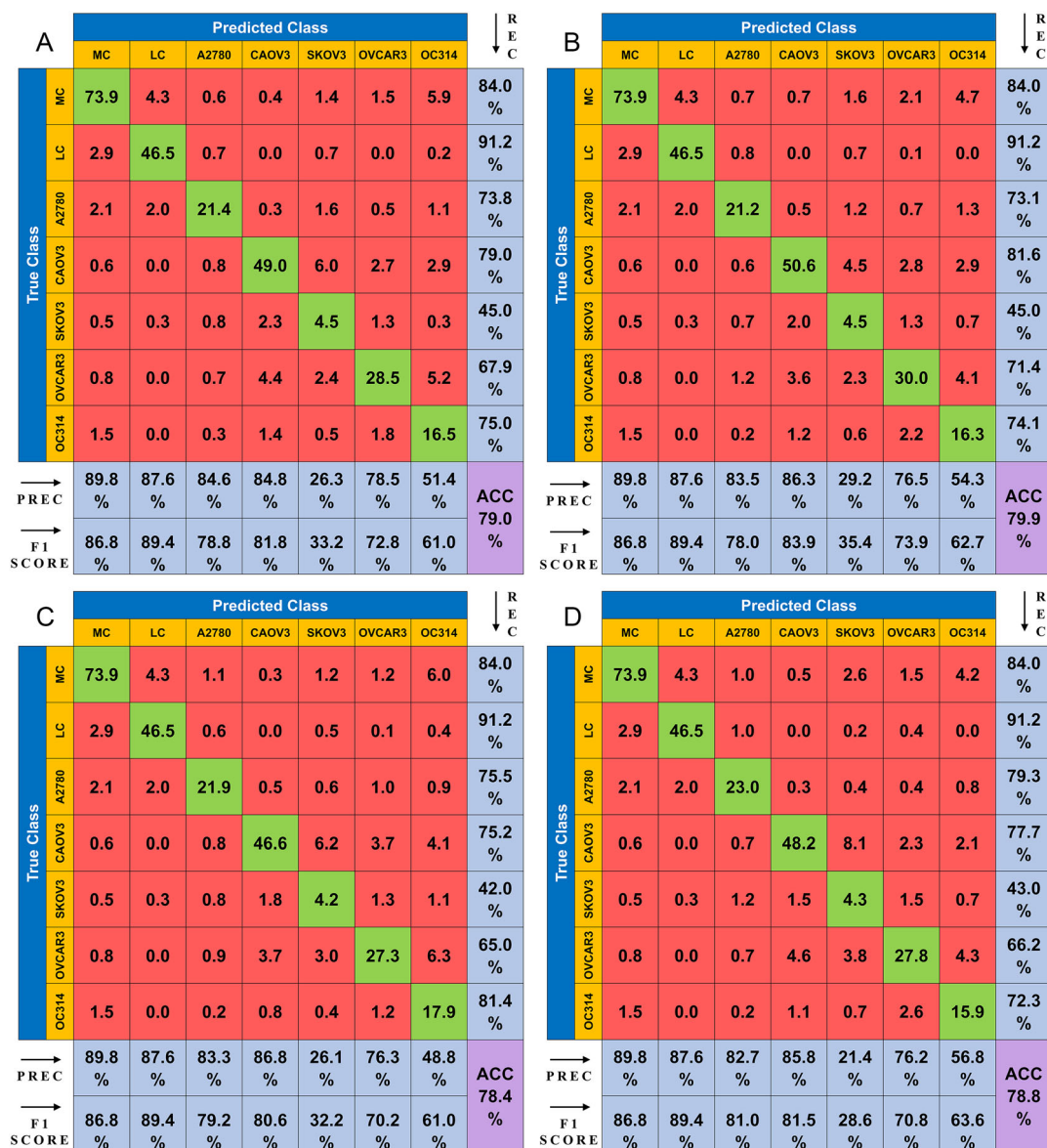


Figure 4. HC confusion matrices related to the cell classification (after max-voting). Details about the several ML model types used in the four HCs are reported in Table S6 (Supporting Information). A) Non-informed HC. B) Histology-informed HC. C) Origin-informed HC. D) Mutation-informed HC. The bottom row contains the precision values. The right column contains the recall values. The bottom right value (violet) is the HC overall accuracy. Non-integer values are due to the average operations among the confusion matrices of the 10 test sets for each HC model.

recovering cells harboring few misclassified QPMs. Of course, MC and LC recalls maintain the same values regardless the HC models, which change only the OCC branch. In particular, following the WBC branch, MCs can be detected with an 84.0% recall (instead of 81.1% without max-voting), while LCs can be detected with a 91.2% recall (instead of 89.3% without max-voting). Instead, the recall enhancement within the OCC branch is different depending on the a priori clinical information. On average, the recall increased by 7.2% in the non-informed HC, by 6.0% in the histology-informed HC, by 6.5% in the origin-informed HC, and by 6.8% in the mutation-informed HC. In order to compare the OCC performance's boosting of the four HC models identification, the average recalls obtained in the QPM case (without max-voting) and in the cell case (with max-voting) were reported in **Figure 5A,B**, respectively. In particular, following the OCC branch in the max-voting case, for detecting. 1) A2780 OCCs, the best result was given by the mutation-informed HC (79.3% recall, 82.7% precision, and 81.0% F1 score); 2) CAOV3 OCCs, the best result was given by the histology-informed HC (81.6% recall, 86.3% precision, and 83.9% F1 score); 3) SKOV3 OCCs, the best result was given by the histology-informed HC (45.0% recall, 29.2% precision, and 35.4% F1 score); 4) OVCAR3 OCCs, the best result was given by the histology-informed (71.4% recall, 76.5% precision, and 73.9% F1 score); 5) OC314 OCCs, the best result was given by the origin-informed HC (81.4% recall, 48.8% precision, and 61.0% F1 score).

For example, to detect A2780, our results proved that the best approach is to use the mutation-informed HC. Instead, the best solution for identification of a metastatic HGSOc was given by the histology-informed HC (OVCAR3) or origin-informed HC (OC314). In general, the histology-informed HC reached the highest accuracy (i.e., 79.9%), while the non-informed HC was never the best solution for classification except for the SKOV3 case for which the non-informed HC provided the same performance as the histology-informed HC after max-voting, even if with a lower precision. Taken together, this means that, in the proposed HC pipeline, exploiting a priori clinical information enhances the classification performance, thus reducing the FN probability.

3.4. An Independent Test to Assess the Identification of Specific OC Subtypes

In order to assess the proposed clinically informed ML approach for the classification of OC subtypes, we performed a further test, independent to those presented in the previous Sections. To this end, we carried out another QPI-FC experiment to collect an additional test set that was as independent as possible from the dataset used so far to train and test HCs. Indeed, we collected a different test set made of 62 CAOV3 cells (with the corresponding 9566 QPMs) and 130 OVCAR3 cells (with the corresponding 29 964 QPMs). As explained in Section 3.3, for both CAOV3 and OVCAR3 OCCs, the best clinically informed HC is the one based on an a priori histology information. For this reason, to assess the classification performance over the independent test set, we selected the histology-informed HC. As explained in Section 3.2, we trained the histology-informed HC with 10 different training sets in order to strengthen the ML performance analysis. Hence, among the 10 trained clinically informed HCs, we selected the two ones having the best recalls over the corresponding original test sets of CAOV3 cells and OVCAR3 cells, respectively. Finally, we employed the trained histology-informed CAOV3-HC and the trained histology-informed OVCAR3-HC to classify OCCs belonging to the independent test sets of CAOV3 cells and OVCAR3 cells, respectively. Results are summarized in **Figure 6**, where we reported the percentage of correctly classified and misclassified CAOV3 cells (see Figure 6A,C) and the percentage of correctly classified and misclassified OVCAR3 cells (see Figure 6B,D). As for the CAOV3 OC cell line, the 51.4% of QPMs is correctly classified (green bar in Figure 6A), meaning that, thanks to the max-voting strategy, the 77.4% of cells is correctly identified (green bar in Figure 6C), which is just the 4.2% lower than the average recall about the original 10 test sets (see Figure 5B). As for the OVCAR3 OC cell line, the 52.8% of QPMs is correctly classified (green bar in Figure 6B), meaning that, thanks to the max-voting strategy, the 68.5% of cells is correctly identified (green bar in Figure 6D), which is just the 2.9% lower than the average recall about the original 10 test sets (see Figure 5B).

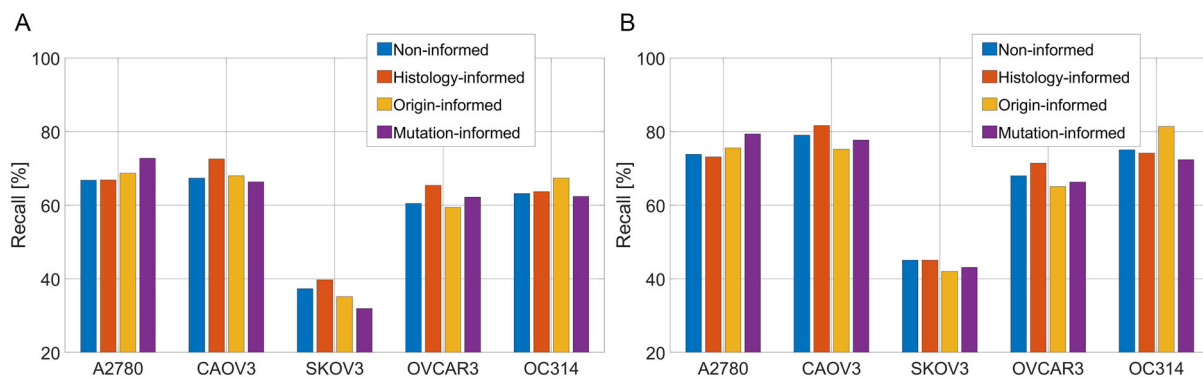


Figure 5. Ovarian cancer cell recall values obtained with each HC model using A) QPM classification (without max-voting) and B) cell classification (with max-voting).

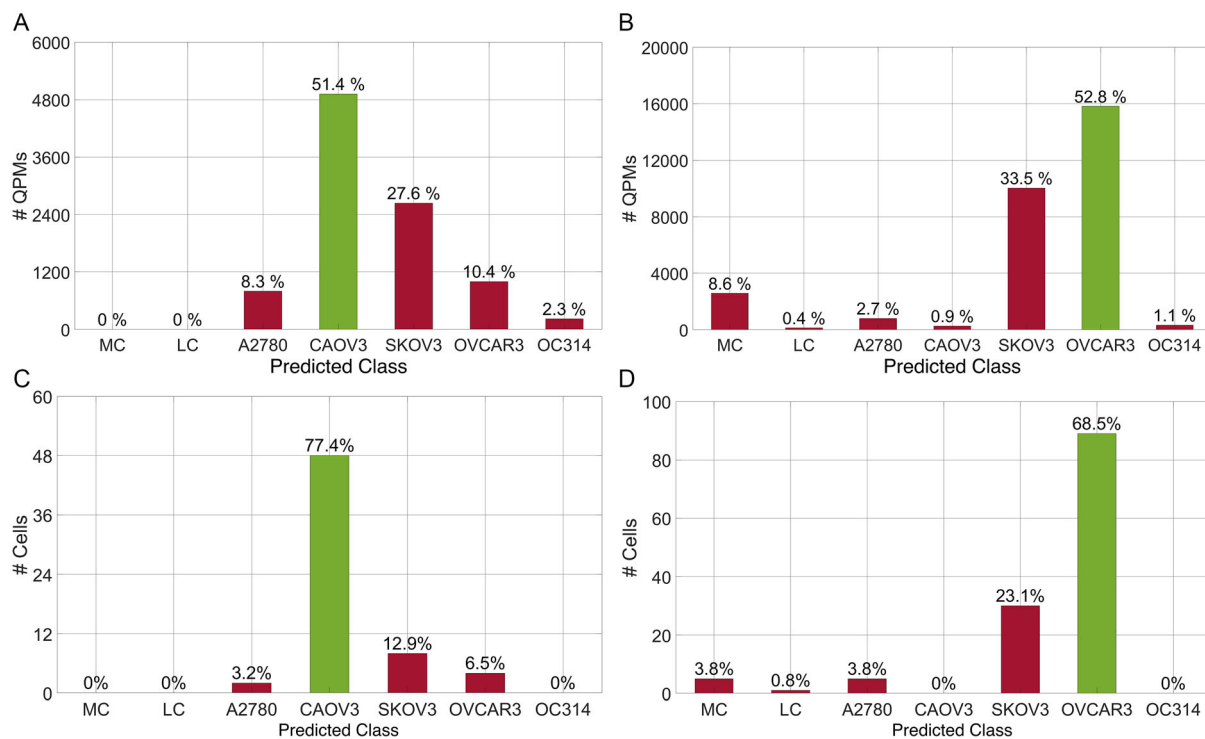


Figure 6. Classification performance of the best histology-informed CAOV3-HC and OVCAR3-HC for the identification of an independent test set of (A,C) CAOV3 OCCs and (B,D) OVCAR3 OCCs, respectively. (A,B) Number of correctly classified QPMs (green) and misclassified QPMs (red). (C,D) Number of correctly classified cells (green) and misclassified cells (red), computed after max-voting. The corresponding percentages with respect to the total number of (A,C) CAOV3 cells and (B,D) OVCAR3 cells are reported at the top of each bar. The green percentage corresponds to the recall value.

4. Discussion and Conclusions

The identification of cancer cells within the blood stream is a very challenging and hot topic nowadays, framed within the so-called liquid biopsy paradigm. QPI offers a unique opportunity to identify OCCs compared to blood cells with similar size and morphology, such as monocytes, thanks to the possibility of measuring RI which acts as an endogenous marker, thus defining a sort of fingerprint of the subcellular structures. Indeed, it is known that OCCs are highly heterogeneous, in terms of molecular and morphological parameters, so it cannot be excluded that some tumor cells share similarities with immunological populations. Therefore, the combination of advanced imaging modalities and AI is crucial to identify OCCs and their subtypes. Indeed, the results reported here confirm that microfluidic flow cytometry coupled to QPI and ML-powered analysis can successfully distinguish cell populations in a label-free manner.^[20]

In particular, we demonstrated the possibility to classify various OCCs from WBCs, which suggests that our approach might successfully handle the intrinsic population heterogeneity and correctly identify OC CTCs in a liquid biopsy of a tested subject during screening. Moreover, we demonstrated the possibility to distinguish, based on label-free QPI, between MCs and LCs. Hence, the proposed method may be capable of simultaneously obtaining complete blood count as well as CTC information, thus further reducing patient management-related time and workload. Specifically, we demonstrated that WBCs can be recognized

with a 91.8% recall while OCCs can be recognized with a 95.3% recall. Then, following the WBC branch, MCs can be detected with an 84.0% recall while LCs can be detected with a 91.2% recall. However, the most significant outcome of our study is reaching better performance for different OC cell lines depending on the available a priori phenotypic/genotypic information. The core of our strategy is the design of a suitable hierarchical scheme based on the type of OC cells identified as target population in a follow-up scenario. Notice that, here we have employed a simple and straightforward ML approach based on handcrafted features extraction and conventional classifiers. However, more sophisticated AI paradigms like deep convolutional neural networks can be easily used in our investigation pipeline to define the best model to be used at each classification level without modifying the HC schemes.

In particular, to detect A2780, results indicated that the best approach is to use the mutation-informed HC (79.3% recall), whereas histology-informed HC is the best option for identification of CAOV3 (91.6% recall), SKOV3 (45% recall) and OVCAR3 cells (71.4% recall). Finally, origin-informed HC is optimal approach for OC314 identification (81.4% recall). To solve each ML classification task, 44 label-free features measured from the QPMs of flowing cells were exploited. All the 44 features were employed to train and test the several ML models. However, we implemented the Relieff algorithm with 10 nearest neighbors^[47] separately for each of the classification levels sketched in Figure 2, in order to rank the predictors according to their

importance in the several classification problems. As summarized in Table S8 (Supporting Information), eccentricity and entropy were often among the highest ranked features, thus suggesting their greater involvement in ML classifications. It is worth noting that the overall performances were also improved thanks to the use of a max-voting strategy, able to exploit multiple views of a cell to further enhance the ML classification at single-cell level through the correlation of multiple inferences. In fact, considering the best recall values of the two WBC cell lines and the five OC cell lines, the sole max-voting allowed increasing the performance on average by 6.6%. Moreover, the combination between the max-voting strategy and the a priori clinical information allowed enhancing the recall values of the OCCs up to 18.3% with respect to the baseline case, that is, the non-informed HC. Of note, the non-informed HC is never the best solution for classification. Indeed, clinically informed HCs achieved higher recall values in all cases, except for SKOV3, possibly due to the lower cell numbers used for the training set. Different OC cell types resulted to be optimally identified by application of different clinically informed HCs, suggesting that a priori knowledge, if appropriately used, improves the reliability in the identification of specific target cells. Interestingly, histology-guided HC resulted to be the best solution for three of the five OC cell lines tested. Although this is a preliminary study, it is reasonable to hypothesize that histological classification of the cells used here is associated with their morphology. Indeed, we observed a distinct cell shape between E/CCOC and HGSOc models when grown in 2D culture (Figure S2, Supporting Information). This difference in morphology, encoded in the biological markers measured by DH microscopy of the rotating cells, possibly explains why histology-informed HC resulted to be most performant. On the contrary, the classification of cells based on their *TP53* genotype is less intuitively correlated to their morphological biomarkers, suggesting caution while making conclusions about our data. Nonetheless, it seems clear that *ad-hoc* use of a priori known phenotype information minimizes the FN errors in detecting the expected OCCs, which is of great importance in the context of using CTC detection for disease and treatment response monitoring.

In summary, the strategy herein proposed can be potentially employed for two different scenarios of label-free liquid biopsy based on ML-powered QPI-FC. The non-informed HC is more suitable for the first diagnosis, when no a priori information is available. In this case, the OCCs could be recognized with respect to the WBCs, then MCs and LCs could be counted simultaneously with the OC phenotyping. Instead, when cancer phenotype is already diagnosed, its follow-up stage can be greatly aided by the proposed clinically informed HCs, since the best a priori information about the specific OC cell line to search is available, thus enhancing the sensitivity of the proposed intelligent decision maker.

Unlike the non-informed HC, the clinically informed HC can also perform intermediate classifications that could be useful in the clinical practice. For example, as summarized in Table 1, OC CTCs characterized by E/CC histology (70.8% recall) can be discriminated from HGSOc (90.2% recall) in the histology-informed HC, or primary tumor-derived OCCs (76.7% recall) can be identified against ascites-derived OCCs (84.7% recall) in the origin-informed HC. Finally, to further assess the

Table 1. Recall values of the intermediate single-cell classification problems (after max-voting) in the different HC models.

HC	Class		
	WBC	OCC	
	91.8%	95.3%	
Histology-informed	Endometrial	Serous	
	70.8%	90.2%	
Origin-informed	Tumor	Ascites	
	76.7%	84.7%	
Mutation-informed	Truncating Mutation	Missense Mutation	Wild-Type
	86.3%	79.1%	79.3%

proposed clinically informed strategy in the follow-up scenario, we performed additional and independent experiments by collecting CAOV3 and OVCAR3 images as new test cases of the trained clinically informed HC models. Remarkably, when tested over the independent test sets, the best clinically informed HCs (i.e., the histology-informed ones for both these two OC cell lines) reached comparable performances with respect to the original test sets, which were a recall of 77.4% and 68.5% in the CAOV3 and OVCAR3 cases, respectively. Moreover, according to the scheme in Figure 2B, in the CAOV3 case, the 100% of cells were correctly recognized as OCCs at the level 1 and the 83.9% of cells were correctly recognized as OC CTCs characterized by E/CC histology at the level 2.2; instead, in the OVCAR3 case, the 95.4% of cells were correctly recognized as OCCs at the level 1 and the 68.5% of cells were correctly recognized as OC CTCs characterized by E/CC histology at the level 2.2. Therefore, the proposed HC scheme proved to be robust under different and independent test cases. In order to further increase the robustness and generalization ability, future investigations will focus on datasets deriving from biological samples with higher complexity level, such as peripheral blood spiked with cancer cells or OC patient-derived material, analogue to what has been reported for renal cell cancer.^[48]

We envision our technology will initially be applied to samples post CTC enrichment. The process is usually preceding most of the CTC detection methods. For example, Parsortix enrichment allows up to 10^5 depletion of nucleated blood cells, with reported residual nucleated cell number <1000 .^[49] Notice that, the high throughput of flow cytometry in immunology and molecular biology approaches is thousands of cells per second. To the best of our knowledge, in the field of liquid biopsy, enrichment, and detection of CTCs by most of the acknowledged methods, including the leading CellSearch+DEPArray workflow, is about 8 h for 7.5 mL of blood.^[50] Our QPI-FC system can collect hundreds of orientations of the same cells by exploiting hydrodynamic forces within the microchannel making all cells rotate during their flow. We have already demonstrated that the potential throughput of our QPI-FC system is tens of full-rotating (360°) cells per second,^[51] thus corresponding to thousands of single-cell images per second taken at different orientations. Therefore, our system may be competitive, considering it brings a significant advantage of not being dependent on cell

labeling. Hence, the proposed analysis opens perspectives for diagnosing the OC cells type in a future scenario of label-free liquid biopsy.^[11] While the significance of liquid biopsy as a screening approach for early cancer diagnosis is still being evaluated, the clinical potential of CTC detection in disease follow-up has already entered clinical practice worldwide for certain cancers, and in OC a higher CTC number has been associated with worse overall survival^[4] and resistance to platinum-based chemotherapy.^[5]

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by project PRIN 2022, Computationally aided Opto-mechano-fluidic pLatform for Label-free intelligent tumor microEnvironment Cell sorting (COLLECT) Prot. 202275PJR, funded by the Italian Ministry of University & Research in the framework of the European Union program Next Generation EU, CUP: B53D23002280006. This work was supported by a project PRIN 2022 PNRR - Flow-cytometry ImaGing by Holographic tomography for predicting TUMor control in Oncology patients treated with Radiotherapy (FIGHT-TUMOR) Prot. P2022ATE2] - funded by the Italian Ministry of University & Research in the framework of the European Union program Next Generation EU, CUP: B53D23023890001. The research leading to these results has received funding partly by Associazione Italiana Ricerca sul Cancro (AIRC) project IG 2019 - ID. 22921 and project IG 2020 - ID. 24494. The authors thank Fondazione Cassa di Risparmio in Bologna (Italy) for the financial support to IK finalized to the acquisition of EVOS M5000. The authors wish to thank also Dr. Danila Del Giudice for her support in samples preparation. Dr. Martina Mugnano contributed to the manuscript during her fellowship with CNR-ISASI, Institute of Applied Sciences and Intelligent Systems "Eduardo Caianiello". Currently, she is affiliated to the Department of Chemical, Materials and Production Engineering, DICMaPI, University of Naples "Federico II", Piazzale Tecchio 80, 80125 Naples, Italy.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Daniele Pirone: Software, Formal Analysis, Writing—Original Draft. **Beatrice Cavina:** Investigation, Writing—Review and Editing. **Daniele Gaetano Sirico:** Investigation. **Martina Mugnano:** Resources. **Vittorio Bianco:** Methodology, Writing—Review & Editing. **Lisa Miccio:** Investigation, Writing—Review & Editing. **Anna Myriam Perrone:** Writing—Review & Editing. **Anna Maria Porcelli:** Resources, Writing—Review & Editing, Funding acquisition. **Giuseppe Gasparre:** Resources, Writing—Review & Editing, Supervision, Funding acquisition. **Ivana Kurelac:** Conceptualization, Investigation, Resources, Writing—Original Draft, Supervision, Funding acquisition. **Pasquale Memmolo:** Conceptualization, Supervision, Methodology, Writing—Original Draft, Funding acquisition. **Pietro Ferraro:** Conceptualization, Supervision, Methodology, Writing—Original Draft, Funding acquisition.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

cancer cells, holographic microscopy, imaging flow cytometry, liquid biopsy, machine learning

Received: May 15, 2024

Revised: July 31, 2024

Published online:

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, *CA Cancer J. Clin.* **2023**, *73*, 17.
- [2] D. B. Asante, L. Calapre, M. Ziman, T. M. Meniawy, E. S. Gray, *Cancer Lett.* **2020**, *468*, 59.
- [3] A. Poveda, S. B. Kaye, R. McCormack, S. Wang, T. Parekh, D. Ricci, C. A. Lebedinsky, J. C. Tercero, P. Zintl, B. J. Monk, *Gynecol. Oncol.* **2011**, *122*, 567.
- [4] E. Obermayr, D. C. Castillo-Tong, D. Pils, P. Speiser, I. Braicu, T. Van Gorp, S. Mahner, J. Sehouli, I. Vergote, R. Zeillinger, *Gynecol. Oncol.* **2013**, *128*, 15.
- [5] N. K. Kim, D. H. Suh, K. Kim, J. H. No, Y. B. Kim, M. Kim, Y. H. Cho, *PLoS One* **2024**, *19*, e0304704.
- [6] S. Lemma, A. M. Perrone, P. De Iaco, G. Gasparre, I. Kurelac, *Am. J. Cancer Res.* **2021**, *11*, 4111.
- [7] U. A. Matulonis, A. K. Sood, L. Fallowfield, B. E. Howitt, J. Sehouli, B. Y. Karlan, *Nat. Rev. Dis. Primers* **2016**, *2*, 1.
- [8] C. Alix-Panabi, ères, K. Pantel, *Nat. Rev. Cancer* **2015**, *14*, 623.
- [9] M. M. Ferreira, V. C. Ramani, S. S. Jeffrey, *Mol. Oncol.* **2016**, *10*, 374.
- [10] S. Ju, C. Chen, J. Zhang, L. Xu, X. Zhang, Z. Li, Y. Chen, J. Zhou, F. Ji, L. Wang, *Biomark. Res.* **2022**, *10*, 1.
- [11] L. Miccio, F. Cimmino, I. Kurelac, M. M. Villone, V. Bianco, P. Memmolo, F. Merola, M. Mugnano, M. Capasso, A. Iolascon, P. L. Maffettone, P. Ferraro, *View* **2020**, *1*, 20200034.
- [12] W. Li, Y. Zhou, Y. Deng, B. L. Khoo, *Cancers* **2022**, *14*, 818.
- [13] A. Muchlińska, J. Smentoch, A. J. Żaczek, N. Bednarz-Knoll, *Cancers* **2022**, *14*, 4178.
- [14] Z. Du, Y. Li, B. Chen, L. Wang, Y. Hu, X. Wang, W. Zhang, X. Yang, *Lab Chip* **2022**, *22*, 3390.
- [15] P. Bankó, S. Y. Lee, V. Nagygyörgy, M. Zrínyi, C. H. Chae, D. H. Cho, A. Telekes, *J. Hematol. Oncol.* **2019**, *12*, 1.
- [16] Y. Park, C. Depeursinge, G. Popescu, *Nat. Photonics* **2018**, *12*, 578.
- [17] K. Lee, K. Kim, J. Jung, J. Heo, S. Cho, S. Lee, G. Chang, J. Jo, H. Park, Y. Park, *Sensors* **2013**, *13*, 4170.
- [18] T. L. Nguyen, S. Pradeep, R. L. Judson-Torres, J. Reed, M. A. Teitell, T. A. Zangle, *ACS Nano* **2022**, *16*, 11516.
- [19] D. Pirone, J. Lim, F. Merola, L. Miccio, M. Mugnano, V. Bianco, F. Cimmino, F. Visconte, A. Montella, M. Capasso, A. Iolascon, P. Memmolo, D. Psaltis, P. Ferraro, *Nat. Photonics* **2022**, *16*, 851.
- [20] C. L. Chen, A. Mahjoubfar, L. C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, B. Jalali, *Sci. Rep.* **2016**, *6*, 21471.
- [21] Q. Ma, D. Xu, *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 303.
- [22] V. Bianco, P. Memmolo, P. Carcagni, F. Merola, M. Paturzo, C. Distanto, P. Ferraro, *Adv. Intell. Syst.* **2020**, *2*, 1900153.
- [23] Z. Guo, X. Lin, Y. Hui, J. Wang, Q. Zhang, F. Kong, *Front. Oncol.* **2022**, *12*, 843879.
- [24] Y. Jo, H. Cho, S. Y. Lee, G. Choi, G. Kim, H. S. Min, Y. Park, *IEEE J. Sel. Top. Quantum Electron.* **2018**, *25*, 1.

- [25] G. Kim, Y. Jo, H. Cho, H. S. Min, Y. Park, *Biosens. Bioelectron.* **2019**, *123*, 69.
- [26] P. Memmolo, G. Aprea, V. Bianco, R. Russo, I. Andolfo, M. Mugnano, F. Merola, L. Miccio, A. Iolascon, P. Ferraro, *Biosens. Bioelectron.* **2022**, *201*, 113945.
- [27] L. L. Zeune, Y. E. Boink, G. van Dalum, A. Nanou, S. de Wit, K. C. Andree, J. F. Swennenhuis, S. A. van Gils, L. W. M. M. Terstappen, C. Brune, *Nat. Mach. Intell.* **2020**, *2*, 124.
- [28] S. Luo, Y. Shi, L. K. Chin, P. E. Hutchinson, Y. Zhang, G. Chierchia, H. Talbot, X. Jiang, T. Bourouina, A. Q. Liu, *Adv. Intell. Syst.* **2021**, *3*, 2100073.
- [29] D. Pirone, A. Montella, D. G. Sirico, M. Mugnano, M. M. Villone, V. Bianco, L. Miccio, A. M. Porcelli, I. Kurelac, M. Capasso, A. Iolascon, P. L. Maffettone, P. Memmolo, P. Ferraro, *Sci. Rep.* **2023**, *13*, 6042.
- [30] A. Cohen, M. Dudaie, I. Barnea, F. Borrelli, J. Běhal, L. Miccio, P. Memmolo, V. Bianco, P. Ferraro, N. T. Shaked, *Adv. Intell. Syst.* **2024**, *6*, 2300433.
- [31] D. Pirone, A. Montella, D. Sirico, M. Mugnano, D. Del Giudice, I. Kurelac, M. Tirelli, A. Iolascon, V. Bianco, P. Memmolo, M. Capasso, L. Miccio, P. Ferraro, *APL Bioeng.* **2023**, *7*, 036118.
- [32] Y. Wang, W. Lin, X. Zhuang, X. Wang, Y. He, L. Li, G. Lyu, *Oncol. Rep.* **2024**, *51*, 1.
- [33] P. Shrestha, B. Poudyal, S. Yadollahi, D. E. Wright, A. V. Gregory, J. D. Warner, P. Korfiatis, I. C. Green, S. L. Rassier, A. Mariani, B. Kim, S. K. Laughlin-Tommaso, T. L. Kline, *Gynecol. Oncol.* **2022**, *166*, 596.
- [34] D. Mikdadi, K. A. O'Connell, P. J. Meacham, M. A. Dugan, M. O. Ojere, T. B. Carlson, J. A. Klenk, *Cancer Biomarkers* **2022**, *33*, 173.
- [35] Cellosaurus resource, <https://www.cellosaurus.org> (accessed: February 2024).
- [36] C. Evangelisti, D. de Biase, I. Kurelac, C. Ceccarelli, H. Prokisch, T. Meitinger, P. Caria, R. Vanni, G. Romeo, G. Tallini, G. Gasparre, E. Bonora, *BMC Cancer* **2015**, *15*, 1.
- [37] D. Pirone, D. Sirico, L. Miccio, V. Bianco, M. Mugnano, P. Ferraro, P. Memmolo, *Lab Chip* **2022**, *22*, 793.
- [38] M. K. Kim, *SPIE Rev.* **2010**, *1*, 018005.
- [39] P. Memmolo, C. Distante, M. Paturzo, A. Finizio, P. Ferraro, B. Javidi, *Opt. Lett.* **2011**, *36*, 1945.
- [40] P. C. Costa, Z. Guang, P. Ledwig, Z. Zhang, S. Neill, J. J. Olson, F. E. Robles, *Biomed. Opt. Express* **2021**, *12*, 1621.
- [41] L. Nanni, S. Brahnam, S. Ghidoni, E. Menegatti, T. Barrier, *PLoS One* **2013**, *8*, e83554.
- [42] F. Borrelli, J. Behal, A. Cohen, L. Miccio, P. Memmolo, I. Kurelac, A. Capozzoli, C. Curcio, A. Liseno, V. Bianco, N. T. Shaked, P. Ferraro, *APL Bioeng.* **2023**, *7*, 026110.
- [43] A. Singh, N. Thakur, A. Sharma, in *2016 3rd Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, March **2016**, pp. 1310–1315.
- [44] C. Beaufort, J. C. Helmijr, A. M. Piskorz, M. Hoogstraat, K. Ruigrok-Ritstier, N. Besselink, M. Murtaza, W. F. J. van Ijcken, A. A. J. Heine, M. Smid, M. J. Koudijs, J. D. Brenton, E. M. J. J. Berns, J. Helleman, *PLoS One* **2014**, *9*, e103988.
- [45] S. Domcke, R. Sinha, D. A. Levine, C. Sander, N. Schultz, *Nat. Commun.* **2013**, *4*, 2126.
- [46] M. S. Anglesio, K. C. Wiegand, N. Melnyk, C. Chow, C. Salamanca, L. M. Prentice, J. Senz, W. Yang, M. A. Spillman, D. R. Cochrane, K. Shumansky, S. P. Shah, S. E. Kalloger, D. G. Huntsman, *PLoS One* **2013**, *8*, e72162.
- [47] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, J. H. Moore, *J. Biomed. Inform.* **2018**, *85*, 189.
- [48] S. Wang, Y. Zhou, X. Qin, S. Nair, X. Huang, Y. Liu, *Sci. Rep.* **2020**, *10*, 12226.
- [49] M. C. Miller, P. S. Robinson, C. Wagner, D. J. O'Shannessy, *Cytometry A* **2018**, *93*, 1234.
- [50] A. J. Rushton, G. Nteliopoulos, J. A. Shaw, R. C. Coombes, *Cancers* **2021**, *13*, 970.
- [51] D. Pirone, M. M. Villone, P. Memmolo, Z. Wang, V. Tkachenko, W. Xiao, L. Che, L. Xin, X. Li, F. Pan, P. Ferraro, P. L. Maffettone, *Opt. Lasers Eng.* **2022**, *158*, 107190.