Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Human-in-the-loop deep learning for scalable detection of circulating tumor cells and cancer-associated fibroblasts by imaging flow cytometry: a dual-center validation study

Jędrzej Świeżewski
Appsilon
Anna Muchlińska
Medical University of Gdańsk
Krystian Zieliński
Appsilon
Aleksandra Parteka
Medical University of Gdańsk
Anna Żaczek
Medical University of Gdańsk
Agnieszka Piastowska-Ciesielska
Medical University of Łódź
Wojciech Ciesielski
Medical University of Łódź
Adam Durczyński
Medical University of Łódź
Rafał Dziadziuszko
Medical University of Gdańsk
Jarosław Kobiela
Medical University of Gdańsk
Elżbieta Senkus
Medical University of Gdańsk
Kevin Miszewski
Medical University of Gdańsk
Marcin Matuszewski
Medical University of Gdańsk
Natalia Bednarz-Knoll
nbk@gumed.edu.pl

Medical University of Gdańsk

Article

Keywords: cancer progression, liquid biopsy, imaging flow cytometry, circulating tumor cells, epithelial-mesenchymal phenotypes, heterogeneity, circulating cancer-associated fibroblasts, peripheral blood, tumor draining-vein blood, diagnostics, digital / computational pathology, deep learning, automation

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11677722/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Human-in-the-loop deep learning for scalable detection of circulating tumor cells and cancer-associated fibroblasts by imaging flow cytometry: a dual-center validation study

Jędrzej Świeżewski 1#*, Anna Muchlińska 2#, Krystian Zieliński 1^, Aleksandra Parteka 2^, Anna Żaczek 2, Agnieszka Piastowska-Ciesielska 3, Wojciech Ciesielski 4, Adam Durczyński 4, Rafał Dziadziuszko 5, Jarosław Kobiela 6, Elżbieta Senkus 5, 7 Kevin Miszewski 8, Marcin Matuszewski 8, Natalia Bednarz-Knoll 2*

#, ^ equal contribution

1 Appsilon, Poland;
2 Laboratory of Translational Oncology, Intercollegiate Faculty of Biotechnology, Medical University of Gdańsk, Poland;
3 BRaIn Laboratories, Medical University of Łódź, Poland;
4 Department of General and Transplant Surgery, Medical University of Łódź, Poland;
5 Department of Oncology and Radiotherapy, Medical University of Gdańsk, Poland;
6 Department of Surgical Oncology, Transplant Surgery and General Surgery, Medical University of Gdańsk, Poland;
7 Breast Unit, University Clinical Center, Gdańsk, Poland;
8 Department of Urology, Medical University of Gdańsk, Poland

* Corresponding author: Natalia Bednarz-Knoll, Laboratory of Translational Oncology, Intercollegiate Faculty of Biotechnology, Medical University of Gdańsk, ul. Dębinki 1, 80-210 Gdańsk, Poland; e-mail: nbk@gumed.edu.pl; Jędrzej Świeżewski, Appsilon, Poland; e-mail: swiezew@gmail.com

Abstract

Circulating tumor cells and cancer-associated fibroblasts (CTCs/cCAFs) are promising biomarkers, but their rarity and heterogeneity make identification labor-intensive, even with standardized technologies such as imaging flow cytometry (imFC). Here, we developed a human-in-the-loop deep-learning workflow addressing this bottleneck by prioritizing candidate cells for expert review.

A model trained on five-channel imFC-images and expert-annotated breast cancer reference CTCs/cCAFs libraries was evaluated in 59 patients with 5 cancer types and 6 healthy donors at two centers/devices.

In peripheral blood, the model detected 443/490 CTCs/cCAFs among 7.8 million events, with 54.8 false positives/100,000 events. Reviewing 50 prioritized candidates/sample reached the predefined threshold in eligible samples, taking 2min 15sec/sample versus an estimated 30min manually. In tumor-draining vein blood, 1,200/1,678 CTCs/cCAFs were detected. Detection of ≥1CTC was associated with lymph node and/or distant metastases (p=0.016).

This workflow may facilitate scalable imFC analysis with expert verification. Multicenter studies are needed to confirm generalizability and clinical relevance.

Key words

cancer progression, liquid biopsy, imaging flow cytometry, circulating tumor cells, epithelial-mesenchymal phenotypes, heterogeneity, circulating cancer-associated fibroblasts, peripheral blood, tumor draining-vein blood, diagnostics, digital / computational pathology, deep learning, automation

Abbreviations

BC indicates breast cancer, cCAF - circulating cancer-associated fibroblast, CRC - colorectal cancer, CTC - circulating tumor cells, FN - false negative, FP - false positive, imFC - imaging flow cytometry, LC - lung cancer, PDAC - pancreatic adenocarcinoma, PB - peripheral blood, RC - renal cancer, TDVB - tumor draining vein blood

Introduction

Circulating tumor cells (CTCs), which disseminate from primary or metastatic tumors into the bloodstream, represent a clinically relevant component of liquid biopsy. Their detection and characterization provide a minimally invasive means of assessing tumor biology through repeated blood sampling and have been associated with prognosis and treatment response across multiple solid tumor entities[1]. Beyond their potential as biomarkers, CTC phenotypic diversity may reflect tumor heterogeneity and metastatic potential that cannot be readily captured by bulk molecular analyses of primary tumors[2–4]. Other circulating cellular components, including circulating cancer-associated fibroblasts (cCAFs) and circulating endothelial cells, may also contribute to metastatic processes and clinical outcomes and can be detected alongside CTCs in multiplex assays[5]. Together, these circulating cell populations may provide complementary information for patient stratification, treatment monitoring, and longitudinal assessment of tumor evolution.

Despite this potential, the clinical implementation of circulating cell-based liquid biopsy remains challenging[6]. CTCs are exceptionally rare, and cCAFs appear to be even less frequent, requiring highly sensitive methods capable of distinguishing these cells from abundant normal blood cells. Established platforms such as CellSearch have demonstrated the clinical value of CTC enumeration. However, their reliance on predefined marker-based strategies may limit the detection of phenotypically heterogeneous CTC populations, particularly cells with low or heterogeneous expression of canonical epithelial markers and non-canonical phenotypes[7]. Methods integrating multiple phenotypic and morphological features at the single-cell level are therefore needed to broaden the detection of CTCs and other tumor-associated circulating cells.

Imaging flow cytometry (imFC) offers an attractive platform for this purpose by combining the high throughput and multimarker capabilities of conventional flow cytometry with high-resolution imaging of individual cells[8]. It enables simultaneous assessment of cellular morphology, fluorescence-marker expression, and intracellular signal distribution, while allowing millions of events to be acquired in a single experiment. Using different inclusion and

exclusion markers, imFC can support the simultaneous detection of distinct circulating cell populations. We previously demonstrated this capability in a one-tube assay for the co-detection of phenotypically diverse CTCs and cCAFs in breast cancer (BC)[5]. However, the high throughput of imFC also creates a major analytical bottleneck: potential rare cells represent only a small fraction of acquired events, and their identification typically requires sequential gating followed by manual inspection of hundreds or thousands of candidate images by experienced operators[5,8]. This process is labor-intensive, time-consuming, and susceptible to inter-operator variability, limiting its scalability and reproducibility.

Deep learning-based computer vision offers an opportunity to address this bottleneck by recognizing complex cellular patterns directly from high-dimensional image data. Such approaches have increasingly transformed image interpretation in radiology, pathology, and cytology, with the potential to reduce manual workload and improve analytical consistency[9–11]. Deep learning has also been investigated for CTC detection, but most published approaches have relied on conventional microscopy or specialized microfluidic imaging platforms and have often been evaluated within a single cancer type, institution, or acquisition setting[12–15] . Their generalizability across heterogeneous CTC phenotypes, tumor entities, clinical centers, blood sources, and imaging instruments remains insufficiently established. Although Bootsma et al.[16] demonstrated generalizability across tumor entities in a large donor cohort, all samples were processed using the same imaging instrument and the analysis focused on epithelial phenotypes. Deep learning-assisted CTC detection in imFC therefore remains comparatively underexplored. In particular, alternative blood sources such as tumor-draining vein blood (TDVB), which may provide higher numbers of CTCs, have not been adequately investigated in this context.

For clinical translation, a deep learning system should demonstrate not only accurate cell classification under controlled conditions but also robustness to biological heterogeneity and reproducibility across independent clinical and technical settings. Importantly, its role may be to support rather than to replace expert assessment by prioritizing candidate cells for review and reducing the number of images requiring manual inspection. Such a human-in-the-loop approach could make high-throughput CTC analysis more efficient while retaining expert verification.

In this study, we aimed to develop and independently validate a deep learning-based workflow for the pre-selection of candidate CTCs and cCAFs in imFC data. The model was trained using

an expert-annotated reference library of CTCs and cCAFs and applied to multimarker immunofluorescence images to identify phenotypically heterogeneous circulating cell populations. Our previously established imFC assay enabled the co-detection of canonical pan-keratin-positive CTCs, putative non-canonical phenotypes, including pan-keratin-negative/vimentin-positive and pan-keratin-negative/vimentin-negative cells, and cCAFs[5]. Here, we evaluated the workflow across multiple solid tumor entities, two clinical centers, and two imFC instruments, including its applicability even to TDVB samples. We also assessed its potential to reduce manual image-review workload while maintaining detection performance across diverse biological and technical settings.

Results

Dataset composition and study design

All samples in the study were processed using a previously described imFC-based one-tube assay incorporating multimarker immunofluorescence staining with inclusion and exclusion markers to identify different CTC phenotypes alongside cCAFs[5,17,18] (Figure 1). The distribution of healthy donors and patient cohorts across the training, validation, and test datasets is presented in Figure 2 and detailed in Table S1.

To assess the generalizability of the developed models, training and validation datasets were restricted to peripheral blood (PB) samples from patients with BC collected and processed at a single center, the University Clinical Center in Gdańsk. The BC cohort used to establish the initial CTC/cCAF reference library has been described previously[5]. The training cohort included samples from three healthy donors, one BC patient without detectable CTCs (CTC-negative), four CTC-positive patients without distant metastases (CTC-positive M0), and three CTC-positive patients with distant metastases (CTC-positive M1). In addition, 91 expert-annotated CTCs and cCAFs from 36 other BC patients, not included in the validation or test cohorts, were incorporated into the training dataset. The validation cohort comprised two donors from each of the four groups: healthy donors, CTC-negative patients, CTC-positive M0 patients, and CTC-positive M1 patients.

The trained models were subsequently evaluated using an independent test dataset (6 healthy donors and 59 cancer patients). The BC component of this cohort included five CTC-negative

patients, seven CTC-positive M0 patients, and six CTC-positive M1 patients. The remaining test samples comprised 50 samples from 41 patients with other malignancies, including lung cancer (LC, n = 15), colorectal cancer (CRC, n = 4), and renal cancer (RC, n = 18), recruited at the University Clinical Center in Gdańsk, as well as pancreatic ductal adenocarcinoma (PDAC, n = 4) patients treated at the Medical University of Łódź. For some patients, tumor-draining vein blood (TDVB) samples were collected in addition to PB samples and analyzed in parallel (n = 11).

Model architecture and direct classification performance

Deep learning models based on several backbone architectures were trained to classify five-channel images of individual cells into two categories: CTCs/cCAFs and normal blood cells. Among the evaluated architectures, ConvNeXt Small achieved the best overall performance and was selected for further analyses. A comparison of the evaluated architectures is provided in Figures S1–S3.

The model demonstrated high detection performance in the independent BC test set, correctly identifying 129 of 141 reference CTCs and cCAFs, with 127 false-positive (FP) predictions among approximately 1.8 million analyzed events (Table 1 and Table S2).

Evaluation across different cancer types

To investigate whether the model could generalize beyond the cancer type represented in the training data, we evaluated its performance in independent test cohorts comprising patients with LC, CRC, and RC. All three tumor entities are carcinomas, although they differ substantially in their molecular characteristics from BC, which was used for model training. The model identified 190 of 216 reference CTCs in LC samples, with 150 FP predictions among approximately 1.1 million analyzed events. In the CRC cohort, one of one reference CTC was detected, with 116 FP predictions among approximately 0.6 million events. In the RC cohort, 111 of 120 reference CTCs were identified, with 3,709 FP predictions among approximately 3.1 million analyzed events (Table 1 and Table S2).

Evaluation across clinical centers and imaging instruments

To assess performance in an independent clinical and technical setting, we analyzed PB samples from four patients with PDAC recruited at the Medical University of Łódź. These samples were collected at a different clinical center, processed by an independent team, and acquired using a different imFC instrument from that used for training and the other test cohorts.

The model detected all 12 reference CTCs in this cohort, generating 189 FP predictions among approximately 1.2 million analyzed events (Table 1 and Table S2). These findings provide preliminary evidence of model applicability across independent clinical and imaging settings, although the small number of patients limits the conclusions that can be drawn.

Evaluation in tumor-draining vein blood

While the training, validation, and other test cohorts comprised PB samples, we additionally evaluated model performance in TDVB samples from patients with PDAC (n = 3), CRC (n = 4), and RC (n = 4). The PDAC samples were collected at the independent clinical center, processed by the independent team, and acquired using the alternative imFC instrument.

Model performance in TDVB was lower than in PB samples, with 1,200 of 1,678 reference CTCs detected. A particularly high number of FP predictions was observed in the RC cohort. Additional procedural filtering and alternative gating strategies were subsequently investigated to reduce FP predictions in RC samples (Figure S5C and Table S3).

Human-in-the-loop workflow and potential reduction in expert review workload

We next investigated the potential application of the model as a candidate-prioritization tool within a human-in-the-loop workflow (Figure 3). In this approach, the model ranks acquired cell images according to their prediction scores, allowing experts to review a predefined number of top-ranked candidates rather than manually inspecting all gated events.

As no definitive threshold for a clinically relevant CTC/cCAF count has been established[19–22], we evaluated candidate detection using thresholds ranging from one to six CTCs/cCAFs per sample and review budgets of up to 100 top-ranked objects per sample. Importantly, this ranking-based analysis was distinct from direct model classification using a fixed prediction threshold of 0.5. Cells classified as negative during direct classification could be recovered among the top-ranked candidates which happened most prominently in the RC cohort, while

correctly classified positive cells could fall outside the selected review budget, which necessarily happened in samples containing more CTCs/cCAFs than the number of images selected for review.

Figure 4A shows the proportion of eligible samples in which the predefined detection threshold was reached at different review budgets. At thresholds of four to six CTCs/cCAFs, all eligible samples reached the predefined detection threshold within the evaluated review budgets. When RC samples were excluded, this performance extended across all investigated thresholds, ranging from one to six CTCs/cCAFs (Figure S5A). The review depth required to reach the threshold of four CTCs/cCAFs in individual eligible samples is presented in Figure 4B.

To assess the practical feasibility of this workflow, two independent experts reviewed the 50 highest-ranked candidate cells per sample from 63 PB test samples, comprising healthy donors and patients with BC, RC, LC, PDAC, and CRC. Inter-observer concordance was 96.6%. At a threshold of four positive events, the mean detection rate across the two experts was 89.7% among samples containing reference CTCs/cCAFs.

The mean time required to review an individual cell was 2.7 s (median: 2.0 s), corresponding to an average review time of approximately 2 min 15 s per sample. By comparison, fully manual, expert-dependent review was estimated to require approximately 30 min per sample and typically involved the assessment of 500–700 gated objects, based on the previous experience of 6 years of work in this field.

CTC phenotype distribution and detection performance

The applied imFC assay enabled originally the assessment of CTCs representing four combinations of keratin (K) and vimentin (V) expression: K+V–, K+V+, K–V+, and K–V–, alongside cCAFs (Figure 1). The distribution of these phenotypes and the corresponding model detection performance are summarized in Table 2.

Although the number of reference cells in some phenotypic categories was limited, the model detected cells across all analyzed CTC phenotypes. Importantly, all 21 reference cCAFs in the test dataset were correctly identified. TDVB samples accounted for 91% of the model's errors despite representing only 24% of the analyzed objects, indicating that the reduced performance in this blood source contributed substantially to the overall detection errors.

Association between CTC detection and clinical characteristics

An exploratory analysis was performed to investigate potential associations between CTC detection and clinical characteristics (Table S4). Patients with at least one detected CTC were more likely to have lymph node involvement and/or distant metastases (N and/or M) than patients in whom no CTCs were detected (n=38 / 69.1% vs. n=7 / 36.8%, respectively; Fisher's exact test, p = 0.016). A similar association was observed at a threshold of at least three CTCs (71.8% vs. 48.6%; p = 0.041), while the association at a threshold of at least four CTCs was borderline significant (72.2% vs. 50.0%; p = 0.050). The direction of the association was consistent across all five CTC thresholds examined. In contrast, no statistically significant associations were observed between CTC detection at any of the tested thresholds and subsequent disease progression (all p ≥ 0.339), although the proportion of patients experiencing progression was numerically higher in the CTC-positive groups.

Discussion

The reliable identification of CTCs and other circulating cancer-associated cells remains challenging because of their rarity and phenotypic heterogeneity. Here, we demonstrate the feasibility of integrating deep learning-assisted candidate cells prioritization into imFC-based detection of CTCs and cCAFs. Rather than replacing expert assessment, the proposed human-in-the-loop workflow ranks candidate cells for manual verification, substantially reducing the number of images requiring review while retaining the expert as the final decision-maker. This approach addresses a major limitation of high-throughput imaging-based assays, in which millions of events may be acquired per sample but only a small fraction represent cells of interest.

In our test dataset, reviewing up to 50 top-ranked candidates per sample enabled all the predefined sample-level detection thresholds to be reached in eligible samples, with the exception of limitations observed in the RC cohort. For the threshold of four positive events, this could be achieved by reviewing as few as 16 cells per sample in the current dataset. Importantly, the mean time required to assess an individual cell was 2.7 s (median 2.0 s), corresponding to approximately 2 min 15 s per sample. This compares with an estimated 30 min for conventional expert-dependent review of hundreds of gated objects. Although this

comparison should be interpreted in light of the different basis for measuring and estimating the two workflows, it illustrates the potential of model-assisted prioritization to reduce manual workload. The expert review step also provides an opportunity to reject false-positive predictions and recognize questionable and/or unusual cellular events that may not be adequately captured by predefined classification criteria. Indeed, during reference-library development, model-assisted blind review prompted reconsideration of some events initially annotated using conventional gating.

An important strength of this study is the evaluation of model performance beyond the conditions represented during training. The model was trained and validated using PB samples from patients with BC and healthy donors collected at a single center, but was subsequently tested across additional carcinoma types, a second clinical center, and a different imFC instrument of the same type. It was also evaluated using TDVB, a blood source not previously investigated in this context. Despite these differences, the model demonstrated the ability to identify candidate CTCs/cCAFs across the evaluated cohorts. These findings provide preliminary evidence of transferability across biological and technical settings. Nevertheless, the relatively small number of patients in some cancer-specific and external-center cohorts warrants cautious interpretation and highlights the need for larger multicenter validation studies.

The present approach also differs from previously reported strategies for automated CTC analysis in terms of imaging platform and the definition of target cells. Most earlier studies based on CellSearch images have used epithelial marker-dependent enrichment and classification. Early feature-based algorithms generated CTC counts with prognostic value comparable to operator assessment[23]. Zeune et al.[12] reported high classification performance using an autoencoding convolutional network, although subsequent studies highlighted the risk of false-positive CTC identification when models are trained predominantly on easily classified events[12,24]. More recently, Husseini-Wüsthoff et al.[25] used self-supervised learning to prioritize CellSearch images for expert review, reducing the number of events requiring manual assessment while recovering a similar number of CTCs [25]. Other approaches have used microfiltration, digital holography, or tumor-specific markers to extend automated detection beyond conventional CellSearch analysis[26,27].

The large-scale SEE-TC study further demonstrated the potential of machine learning-based CTC analysis across multiple cancer entities, reporting associations between longitudinal CTC

counts and overall survival[16]. However, its samples were processed centrally using a single imaging platform, and CTC identification relied on tumor-specific inclusion markers. In contrast, our study used imFC images obtained from a multimarker one-tube assay performed directly on patient blood samples, enabling the assessment of several putative CTC phenotypes alongside cCAFs. We also evaluated performance across independent clinical and imaging settings and included TDVB samples. Thus, our work extends previous efforts by investigating deep learning-assisted candidate prioritization in a high-throughput imaging platform and across a broader range of circulating cell phenotypes and analytical conditions.

The ability to detect different CTC phenotypes is particularly relevant because conventional epithelial marker-based approaches may not capture the full diversity of CTCs. In our assay, candidate CTCs included K+V–, K+V+, K–V+, and K–V– phenotypes, alongside cCAFs. Importantly, the model was trained to identify a combined CTC/cCAF-positive class rather than to distinguish these individual phenotypes. Nevertheless, the model detected reference cells across the investigated phenotypic categories. In an additional analysis restricted to epithelial-positive phenotypes, the models trained on the former task also demonstrated good performance, suggesting that the proposed approach can be applied to a more narrowly defined target population without retraining. This comparison should, however, be interpreted as an alternative evaluation of the existing models rather than an independently optimized epithelial-specific classifier (see Figure S4).

An additional finding of potential clinical relevance was the association between CTC detection and lymph node and/or distant metastatic involvement. Patients with at least one detected CTC were more likely to have lymph node and distant organs involvement than those without detectable CTCs (69.1% vs. 36.8%; $p = 0.016$). A similar direction of association was observed across the different CTC thresholds. This finding is consistent with the biological role of CTCs in hematogenous dissemination and suggests that even low CTC counts may carry clinically relevant information. However, no statistically significant association was observed between CTC detection and subsequent disease progression during the available follow-up. Given the limited cohort size, heterogeneity of the included malignancies, and multiple thresholds examined, these clinical analyses should be considered exploratory and hypothesis-generating. Larger prospective studies with longer follow-up are needed to determine whether CTC detection provides independent prognostic information.

Study limitations

Several limitations should be considered when interpreting these findings. First, although the model demonstrated high detection performance, it did not identify all reference CTCs/cCAFs. Some cells, including those occurring as single events in a sample, were missed. The clinical relevance of detecting individual CTCs remains uncertain[23], and our results do not establish an optimal clinically meaningful detection threshold. Notably, missed cells were not restricted to a particular phenotype and were even more frequent among K+V– cells (Table 2).

Second, three healthy donors aged over 70 years each had one putative CTC identified during iterative model-assisted expert review. These cells, displaying K–V– or K–V+ phenotypes, had not been identified during conventional imFC gating-based analysis performed by the experts. As the subsequent clinical follow-up of those healthy donors was unavailable, the biological significance of those cells remains uncertain. This observation highlights the need for orthogonal molecular or genomic validation, as well as additional morphological criteria, to corroborate the tumor origin of non-canonical CTC-like events, particularly those lacking epithelial marker expression.

Third, model performance was less consistent in RC samples, which generated a relatively high number of false-positive predictions. We therefore investigated whether more stringent gating could improve performance in four RC samples (RC22, RC53, RC57, and RC59). Restricting the analysis to DAPI-positive, CD45/CD31-negative events reduced the number of objects processed by the model 26-fold, from approximately 1.5 million to 56,135, and reduced false-positive predictions 18-fold, to 677, without an apparent loss of CTC/cCAF detection in the PB samples. However, this approach increased the number of missed CTCs in TDVB, indicating that stricter gating may have different consequences depending on the blood source. An alternative strategy, in which predictions were calculated on gated events but normalized using statistics from the full sample, produced more favorable results, with an approximately twofold reduction in false positives without additional missed CTCs. Nevertheless, some clustered TDVB-derived CTCs were excluded by the gate. Additional procedural filtering based on foreground-structure similarity between nuclear and vimentin channels also improved model performance in RC samples (Figure S5C). These findings suggest that further optimization of pre-analytical filtering may help reduce false-positive predictions, but should be carefully balanced against the risk of losing rare target cells.

Fourth, although approximately 10.5 million events were analyzed, the number of patients was substantially lower than in large-scale studies such as SEE-TC[16]. The small size of some cancer-specific cohorts, particularly the external-center PDAC cohort, limits the strength of conclusions regarding generalizability. Furthermore, the clinical analysis was exploratory, and the available follow-up was insufficient to establish associations with disease progression or other long-term outcomes.

Finally, the model was based on five imaging channels and did not include α-SMA, which was used in the assay to identify cCAFs. Although this channel was not required to identify the combined CTC/cCAF-positive class, its inclusion would be necessary to distinguish CTCs from cCAFs automatically. Extending the model to incorporate α-SMA and other relevant markers may therefore enable more detailed classification of individual circulating cell subtypes.

Conclusions and future directions

In conclusion, this study demonstrates the feasibility of a deep learning-assisted, human-in-the-loop workflow for the pre-selection of candidate CTCs and cCAFs in imFC-derived datasets. By combining multimarker image analysis with expert verification, the approach reduces the number of events requiring manual inspection while enabling the detection of phenotypically heterogeneous circulating cell populations. Evaluation across multiple carcinoma types, two clinical centers, different imaging instruments, and TDVB samples provides preliminary evidence of its applicability beyond the conditions represented during model development.

Further studies are required before broader clinical implementation can be considered. Larger prospective multicenter cohorts will be needed to confirm generalizability and establish cancer-specific performance, particularly across different blood sources. Additional optimization of candidate selection and pre-analytical filtering should aim to reduce false-positive predictions without compromising sensitivity. Orthogonal molecular or genomic characterization will also be important to establish the biological identity of non-canonical CTC phenotypes. Finally, studies with longer follow-up and established clinical endpoints are needed to determine whether CTC abundance and phenotype-resolved CTC/cCAF profiling provide independent prognostic or treatment-response information. Overall, the proposed

workflow provides a foundation for scalable, expert-supervised analysis of heterogeneous circulating tumor-related cells.

Methods

Patients

The study population comprised healthy individuals and cancer patients from a previously described cohort [5] and participants newly recruited for the ongoing studies, which served as the training, validation and test cohorts (Fig. 1). Patients were treated by systemic treatment or surgery according to national recommendations. Clinico-pathological parameters of the patients, and their inclusion in the study sub-cohorts are summarized in Table S1. Written informed consent was obtained from all participants prior to study enrollment. The study protocols were approved by the relevant local ethics committees (protocol no.: NKBBN/286/2018, NKBBN/748/2019–2020, NKBBN/434/2017, KB/616/2026, and RNN/50/22/KE).

Blood sample collection

PB was collected preoperatively or before initiation of systemic treatment from all recruited patients into EDTA tubes at the vol. 7.5 mL. First, 3 mL of blood was discarded to avoid the risk of contamination with skin cells and endothelial cells, which may occur during puncture. TDVB samples were collected intraoperatively at the maximum feasible volume (usually 2–3 mL), before tumor removal, from veins located downstream to the tumor. All blood samples were processed as described[5] as soon as possible after collection, with most samples processed within 3 hours. The (peripheral) blood mononuclear cell ((P)BMC) fraction, potentially containing CTCs, was isolated using density gradient centrifugation: 6 mL of blood was layered onto 4 mL of Ficoll-Paque™ PLUS, Cytiva, cat. no. 17144003, and centrifuged at 400 × g for 30 min at room temperature without acceleration or deceleration). The (P)BMC fraction was collected, mixed with paraformaldehyde (PFA) solution to reach a final concentration of 4% PFA, and stored at –80°C for further analysis.

Immunofluorescence staining protocol

Immunofluorescence staining aiming to co-detect CTCs and cCAFs was performed as previously described[5]. Briefly, PFA-fixed cells were washed with 1× PBS and incubated for 30 min at 4 °C with an antibody cocktail prepared in 1× Perm/Wash Buffer (BD Biosciences). The staining panel included antibodies against epithelial cells marker (pan-keratins, i.e., K, AE1/AE3 and C11 clones, AF488-conjugated, Thermo Fisher Scientific), mesenchymal cells marker (vimentin, i.e., V, D21H3 clone, AF647-conjugated, Cell Signaling), fibroblasts marker (α-smooth muscle actin, i.e., α-SMA, 1A4 clone, PE-conjugated, R&D Systems), fibroblasts / mesenchymal stem cells marker (CD29, TS2/16 clone, SuperBright600-conjugated, Thermo Fisher Scientific), leukocytes marker (CD45, REA747 clone, APC-Vio770-conjugated, Miltenyi Biotec), and endothelial cells marker (CD31, WM59 clone, APC-Cy7-conjugated, BioLegend). Following incubation, cells were washed with 1× PBS, resuspended in 30 µL of 1× PBS, and counterstained with DAPI (BD Biosciences; 1 µg/mL). Of note, due to the early stage of assay development at the time of analysis, α-SMA and CD29 were not included in the antibody cocktail and thus not analyzed in LC. Samples were subsequently analyzed using the Amnis® ImageStream® X Mk II (Cytek® Biosciences).

Imaging flow cytometry acquisition

Stained samples were acquired using the Amnis® ImageStream® X Mk II (Cytek® Biosciences), equipped with 405 nm, 488 nm, and 642 nm lasers and INSPIRE™ software (version 201.2.0.7). Laser powers were set to 100 mW, 100 mW, and 150 mW, respectively. Cells were imaged at 40× magnification at low flow speed. Fluorescence compensation was established using single-stained controls, as previously described[5], and the optimized instrument settings were kept constant across all clinical samples. The event selection strategy differed depending on cancer type. For BC, LC, and CRC samples, DAPI+/CD45–/CD31– events were collected directly during acquisition in INSPIRE™, whereas for RC and PDAC samples, all DAPI-positive events were acquired, followed by CD45–/CD31– gating in IDEAS software (version 6.4.9.0; Cytek® Biosciences). Images of all tumor entities were acquired using the full staining protocol, with 9 channels except LC, which was acquired with limited protocol, using 7 channels (covering the 5 channels set used in the machine learning analysis of the current study).

Expert annotation and creation of the CTC reference library

Within the preselected DAPI+/CD45–/CD31– population, events positive for K and/or V were manually inspected by experts in brightfield and all fluorescence channels[5] . Putative CTCs were required to display intact cellular and nuclear morphology and a distinct, cell-associated K and/or V fluorescence pattern, without evidence of autofluorescence, fluorescence spillover, or non-specific staining. Events with granulocyte-like or apoptotic nuclear morphology, insufficient image focus, or other features inconsistent with intact tumor cells were excluded. cCAFs were defined as a-SMA- and/or CD29-positive cells. Experts were blinded to clinico-pathological data.

For the previously described BC cohort[5], all expert-annotated CTC events from each CTC-positive patient were exported as a separate IDEAS file, and subsequently merged to generate a combined BC CTC reference library containing only confirmed CTC events of different phenotypes, alongside cCAFs. During this process, the association between each individual CTC or cCAF and the patient of origin was retained, enabling patient-level tracking throughout subsequent analyses.

Several rounds of blind review of cells have been conducted (following the human-in-the-loop principle utilised in[25], presenting experts with only the images of the cells. The selection of cells for review was dictated by preliminary machine learning models trained on the initial labels for this purpose. This enabled an improvement of the labels based only on gating. In particular, cells without a CD45/CD31 signal, but with a bright artifact in the channel misaligned with the cell were captured this way.

Image preprocessing

Each donor sample has been exported to individual 16-bit tiff images without padding or masking using the IDEAS software. While more auxiliary channels were collected, only five channels were used for the development of the machine learning model (Bright field, DAPI, CD45/CD31, K and V). The total number of single-channel images used in the machine learning analysis was 52,669,055.

Development of the machine learning models

Models were implemented in PyTorch and trained on a single NVIDIA T4 GPU. Images were first centre-cropped or padded to a fixed 80 x 80 pixels, padding by edge replication. Images were then intensity-normalized per donor: per-channel clipping to fixed cutoffs followed by per-donor standardization, applied identically to padded and measured pixels. Cell sets not originating directly from donors, which enrich the training set with CTCs, were normalized using the average training-set statistics. Prior to normalization, global channel-specific, histogram-based intensity cutoffs have been applied.

During training, each cell image was augmented on-the-fly with random horizontal and vertical flips, random 90° rotations, random rotation (range ±45°) and random shear (range ±20° in x and y), each applied independently with probability 0.5. All transformations were geometric and applied identically across the five channels of a given cell; border pixels were extended by replication. No augmentation was applied at validation or test inference.

The primary model architecture used in the study was a classifier based on a Convnext Small[28] architecture (selected for its relatively modest architecture size compared to its benchmark performance) pretrained on ImageNet and adapted to five-channel input by averaging the pretrained RGB stem weights and replicating them across input channels, with a single-logit classification head. All layers were fine-tuned for 30 epochs using AdamW optimizer (weight decay 0.01, batch size 32) with a one-cycle learning rate schedule peaking at 0.00001. A binary cross-entropy loss was used, weighting positive samples (CTCs and cCAFs) by a factor of 10 to account for their rarity and importance. The final model was selected by minimum validation loss, which occurred at epoch 18. Per-sample performance of the final model on the training, validation and test sets is summarised in Table 1, reported in detail in Table S2 and illustrated in Figure 4.

Model ablation was performed, by training different models, with a training protocol mirroring the one described above. Each time the model with the lowest loss on the validation set has been picked from a training run. From the Convnext family[28] we trained the Small and Base model sizes; ConvNextv2-FCMAE[29] has been selected with weights pretrained using the masked auto-encoder setup; ResNet-50[30] and EfficientNetB0[31] have been trained to compare

to models often used in the literature for fluorescent microscopy; DINOv2[32] was trained as a small-size representative of the transformer family of models. See Figures S1-S3 for the performance comparison of the different models.

Reporting guidelines

The study was reported with reference to relevant EQUATOR Network guidelines, including the STROBE statement for observational research, the BRISQ recommendations for biospecimen reporting, and the CLAIM 2024 checklist for artificial intelligence in medical imaging studies

Ethics declarations

The patients' recruitment and clinical samples collection were conducted in accordance with the Declaration of Helsinki and approved by the Independent Bioethics Committee for Scientific Research at Medical University of Gdańsk (protocol no.: NKBBN/286/2018, NKBBN/748/2019–2020, NKBBN/434/2017 and KB/616/2026) as well as Independent Bioethics Committee for Scientific Research at Medical University fo Łódź (protocol no. RNN/50/22/KE).

Funding

Publication was funded from the State Budget under the programme of the Minister of Education and Science entitled 'Science for Society II', project No. NdS-II/SP/0398/2023/01, amount of funding: 1 000 000,00 PLN, total project value: 1 000 000,00 PLN

Data availability

Data will be available from the corresponding author upon reasonable request. Further images of representative CTCs of different phenotypes, circulating fibroblasts and leukocytes are presented on the CTC Atlas (www.ctcatlas.org).

Code availability

Code will be available from the corresponding author upon reasonable request.

Authors contribution

Conceptualization: JŚ, NBK; Methodology: JŚ, AM, KZ, AP, NBK; Investigation: AM, AP, NBK; Data curation: JŚ, KZ, AM, AP; Software: JŚ, KZ; Formal analysis: JŚ, NBK; Resources: AZ, APC, WC, AD, RD, ES, JK, KM, MM; Supervision: JŚ, NBK; Project administration: NBK; Funding acquisition: NBK; Writing – original draft: JŚ, AM, AP, NBK; Writing – review and editing: all authors.

Acknowledgements

We thank all patients and healthy donors who agreed to donate blood. We thank Dr Beata Pieczyńska-Uziębło, Dr Małgorzata Siatkowska, Dr Cecilia Panek and Michał Maj for blood processing, immunofluorescent stainings and imFC acquisition, as well as Katarzyna Wnuk, Natalia Literska, Barbara Bollin-Matysiak, and Grażyna Suchodolska for clinical data curation. Figures preparation was supported using BioRender under the license for the Medical University of Gdańsk. Please, find further related data on the CTC Atlas: www.ctcatlas.org

Tables

Table 1 Summary of direct model performance on test set
Performance of the Convnext Small model on subsets of the test study is presented. The detected / eligible samples refers to the model's performance on the task of identifying at least 4 CTCs/CAFs in samples with at least 4 positive cells within the review budget of 50.

Test subset	Total objects	Found / total positive	Recall [%]	95% CI (Clopper-Pearson) [%]	False positives per 100,000 cells	Detected / eligible samples
BC / PB	1 822 825	129 / 141	91,50	85.6-95.5	7	6 / 6

LC / PB	1 112 550	190 / 216	88,00	82.9-92.0	13,5	10 / 10
RC / PB	3 101 241	111 / 120	92,50	86.2-96.5	119,6	2 / 2
PDAC / second center / PB	1 245 916	12 / 12	100,00	73.5-100.0	15,2	2 / 2
CRC / PB	550 509	1 / 1	100,00	2.5-100.0	21,1	–
all PB	7 833 041	443 / 490	90,40	87.4-92.9	54,8	20 / 20
TDVB / (PDAC + CRC + RC)	2 425 784	1200 / 1678	71,50	69.3-73.7	577,1	9 / 9

Table 2 Model performance by CTC subtype

The table summarizes the number of cells from the different populations of CTCs and CAFs per split as well as the number of false negative predictions of the Convnext Small model per population. The errors of the model are relatively evenly distributed, with notable perfect performance on the CAFs.

		CTC population					
	Split	**All cells in split (n)**	**K+V+ (n)**	**K+V- (n)**	**K-V+ (n)**	**K-V- (n)**	**CAF (n)**
All cells	train	222 220	22	126	165	143	14
	val	52 766	29	14	6	2	17
	test	10 258 825	831	1 058	208	50	21
False negative	train	222 220	0	1	0	4	0
	val	52 766	2	0	0	0	0
	test	10 258 825	176	318	20	11	0

Table S1 Patient characteristics and machine learning splits

The table summarizes samples from patients used in the study, presented together with their clinical labels metadata and the splits used for training and evaluating machine learning models.

Split	Patient nr	Status	Entity	Clinical centre	Type of blood sample	Age	Sex	Clinical parameters	CTC/cCAF status (expert)	CTC/cCAF status (model + expert)
train	HD1	healthy	HD	Gdańsk	PB	34	female	-	neg	neg
train	HD5	healthy	HD	Gdańsk	PB	52	female	-	neg	neg
train	HD6	healthy	HD	Gdańsk	PB	77	female	-	neg	pos
train	BC8	patient	BC	Gdańsk	PB	55	female	cT3N3	pos	pos
train	BC34	patient	BC	Gdańsk	PB	61	female	M1	pos	pos
train	BC52	patient	BC	Gdańsk	PB	60	female	M1	neg	pos
train	BC80	patient	BC	Gdańsk	PB	39	female	pT2N0	neg	neg
train	BC84	patient	BC	Gdańsk	PB	57	female	cT2N1	pos	pos
train	BC122	patient	BC	Gdańsk	PB	42	female	cT4bN1	pos	pos
train	BC128	patient	BC	Gdańsk	PB	69	female	M1	pos	pos
train	BC200	patient	BC	Gdańsk	PB	39	female	cT2N0	pos	pos
train	BCLibrary1	library	BC	Gdańsk	PB	-	female	-	library	library
train	BCLibrary2	library	BC	Gdańsk	PB	-	female	-	library	library
val	HD3	healthy	HD	Gdańsk	PB	57	female	-	neg	neg
val	HD4	healthy	HD	Gdańsk	PB	39	female	-	neg	neg
val	BC35	patient	BC	Gdańsk	PB	32	female	cT2N0	neg	neg
val	BC45	patient	BC	Gdańsk	PB	55	female	cT4dN1	pos	pos
val	BC49	patient	BC	Gdańsk	PB	63	female	M1	pos	pos
val	BC60	patient	BC	Gdańsk	PB	67	female	cT2N0	neg	neg
val	BC145	patient	BC	Gdańsk	PB	66	female	M1	pos	pos
val	BC151	patient	BC	Gdańsk	PB	44	female	cT2N1	pos	pos
test	HD2	healthy	HD	Gdańsk	PB	68	male	-	neg	neg
test	HD21	healthy	HD	Gdańsk	PB	67	male	-	neg	neg
test	HD22	healthy	HD	Gdańsk	PB	75	male	-	neg	pos
test	HD24	healthy	HD	Gdańsk	PB	47	male	-	neg	neg
test	HD25	healthy	HD	Gdańsk	PB	62	male	-	neg	neg
test	HD26	healthy	HD	Gdańsk	PB	70	male	-	neg	pos
test	BC5	patient	BC	Gdańsk	PB	63	female	M1	pos	pos
test	BC21	patient	BC	Gdańsk	PB	52	female	M1	pos	pos
test	BC25	patient	BC	Gdańsk	PB	62	female	M1	pos	pos
test	BC30	patient	BC	Gdańsk	PB	64	female	cT4dN1	neg	neg
test	BC31	patient	BC	Gdańsk	PB	37	female	cT2N1	pos	pos
test	BC76	patient	BC	Gdańsk	PB	42	female	M1	pos	pos
test	BC77	patient	BC	Gdańsk	PB	52	female	cT1N1	pos	pos
test	BC91	patient	BC	Gdańsk	PB	61	female	cT2N0	neg	neg
test	BC104	patient	BC	Gdańsk	PB	36	female	cT1cN1	neg	neg
test	BC115	patient	BC	Gdańsk	PB	57	female	cT4N2	pos	pos
test	BC125	patient	BC	Gdańsk	PB	62	female	cT4cN1	neg	neg
test	BC132	patient	BC	Gdańsk	PB	77	female	cT4N1	pos	pos

test	BC136	patient	BC	Gdańsk	PB	56	female	cT1cN1	neg	neg
test	BC146	patient	BC	Gdańsk	PB	64	female	cT2N0	pos	pos
test	BC197	patient	BC	Gdańsk	PB	56	female	cT4bN1	pos	pos
test	BC199	patient	BC	Gdańsk	PB	50	female	cT2N1	pos	pos
test	BC219	patient	BC	Gdańsk	PB	53	female	cT2N1	pos	pos
test	BC220	patient	BC	Gdańsk	PB	69	female	cT2N0M1	pos	pos
test	LC4	patient	LC	Gdańsk	PB	54	male	T3N2M0	pos	pos
test	LC5	patient	LC	Gdańsk	PB	53	female	T2aN2M1a	pos	pos
test	LC6	patient	LC	Gdańsk	PB	73	male	T4N0M1c	pos	pos
test	LC9	patient	LC	Gdańsk	PB	42	female	T2aNxMx	neg	neg
test	LC13	patient	LC	Gdańsk	PB	64	female	T4N0M1b	pos	pos
test	LC14	patient	LC	Gdańsk	PB	68	female	T4N3M1b	pos	pos
test	LC15	patient	LC	Gdańsk	PB	43	male	T3N3M1b	neg	neg
test	LC16	patient	LC	Gdańsk	PB	56	male	T4N2M1	neg	pos
test	LC21	patient	LC	Gdańsk	PB	63	male	T4N0M1b	neg	pos
test	LC22	patient	LC	Gdańsk	PB	66	male	M1	pos	pos
test	LC23	patient	LC	Gdańsk	PB	49	female	M1	neg	pos
test	LC24	patient	LC	Gdańsk	PB	59	male	T4N1M1b	pos	pos
test	LC25	patient	LC	Gdańsk	PB	69	male	T2aN2M1b	neg	pos
test	LC27	patient	LC	Gdańsk	PB	71	female	T1N2M1b	neg	pos
test	LC28	patient	LC	Gdańsk	PB	74	female	T3N0M1b	neg	pos
test	PDAC3	patient	PDAC	Łódź	PB	69	female	pT1cN0M0	neg	neg
test	PDAC4	patient	PDAC	Łódź	PB	62	female	T4NxM0	pos	pos
test	PDAC7	patient	PDAC	Łódź	PB	68	male	pT3N2M0	pos	pos
test	PDAC12	patient	PDAC	Łódź	PB	61	male	T4NxM0	neg	neg
test	PDAC3	patient	PDAC	Łódź	TDVB	69	female	pT1cN0M0	pos	pos
test	PDAC4	patient	PDAC	Łódź	TDVB	62	female	T4NxM0	neg	neg
test	PDAC12	patient	PDAC	Łódź	TDVB	61	male	T4NxM0	pos	pos
test	CRC1	patient	CRC	Gdańsk	PB	71	male	pT3N2a	neg	neg
test	CRC2	patient	CRC	Gdańsk	PB	78	female	pT3bN0M0	neg	neg
test	CRC3	patient	CRC	Gdańsk	PB	73	male	pT3N0M0	neg	neg
test	CRC4	patient	CRC	Gdańsk	PB	74	male	pT2N0M0	neg	pos
test	CRC1	patient	CRC	Gdańsk	TDVB	71	male	pT3N2a	pos	pos
test	CRC2	patient	CRC	Gdańsk	TDVB	78	female	pT3bN0M0	pos	pos
test	CRC3	patient	CRC	Gdańsk	TDVB	73	male	pT3N0M0	pos	pos
test	CRC4	patient	CRC	Gdańsk	TDVB	74	male	pT2N0M0	neg	neg
test	RC13	patient	RC	Gdańsk	PB	48	male	pT3aNx	pos	pos
test	RC22	patient	RC	Gdańsk	PB	72	female	pT3aN0M0	pos	neg
test	RC24	patient	RC	Gdańsk	PB	66	male	pT3aN0M0	pos	pos
test	RC29	patient	RC	Gdańsk	PB	64	male	pT1bN0M0	pos	pos
test	RC32	patient	RC	Gdańsk	PB	77	male	pT1bN0M0	neg	neg
test	RC33	patient	RC	Gdańsk	PB	77	female	pT1aN0M0	neg	neg
test	RC35	patient	RC	Gdańsk	PB	68	male	pT3aN0M0	neg	neg
test	RC36	patient	RC	Gdańsk	PB	71	female	pT3aNxM0	pos	pos
test	RC37	patient	RC	Gdańsk	PB	78	female	pT2bNxM0	pos	pos
test	RC42	patient	RC	Gdańsk	PB	46	male	M1	neg	neg
test	RC43	patient	RC	Gdańsk	PB	69	female	pT3aNxM0	pos	pos
test	RC53	patient	RC	Gdańsk	PB	69	male	pT1aN0M0	pos	pos

test	RC57	patient	RC	Gdańsk	PB	65	female	pT1aN0M0	pos	pos
test	RC59	patient	RC	Gdańsk	PB	69	male	pT1bN0M0	pos	pos
test	RC61	patient	RC	Gdańsk	PB	70	male	pT3aN0M0	pos	pos
test	RC62	patient	RC	Gdańsk	PB	53	female	pT1bNxM0	pos	neg
test	RC13	patient	RC	Gdańsk	TDVB	48	male	pT3aNx	pos	pos
test	RC14	patient	RC	Gdańsk	TDVB	76	male	pT3aN0M0	pos	pos
test	RC22	patient	RC	Gdańsk	TDVB	72	female	pT3aN0M0	pos	pos
test	RC25	patient	RC	Gdańsk	TDVB	73	male	pT3aN1M1	pos	pos

Table S2 Model performance per sample

Detailed performance of the Convnext Small model on the samples used in the study is presented.

Split	Patient nr	Objects (n)	CTCs (n)	TP (n)	FP (n)	TN (n)	FN (n)	Rank 4th	Check50 at rank 4th
train	HD1	10120	0	0	1	10119	0	-	-
train	HD5	3088	0	0	2	3086	0	-	-
train	HD6	6197	1	1	1	6195	0	-	-
train	BC8	24503	67	67	38	24398	0	4	detected
train	BC34	2035	21	21	4	2010	0	4	detected
train	BC52	45233	1	1	3	45229	0	-	-
train	BC80	20460	0	0	3	20457	0	-	-
train	BC84	41087	120	115	39	40928	5	4	detected
train	BC122	33402	152	152	55	33195	0	4	detected
train	BC128	31603	2	2	2	31599	0	-	-
train	BC200	4396	15	15	4	4377	0	4	detected
train	BCLibrary1	44	40	40	0	4	0	4	detected
train	BCLibrary2	52	51	51	0	1	0	4	detected
val	HD3	1736	0	0	1	1735	0	-	-
val	HD4	2244	0	0	2	2242	0	-	-
val	BC35	13949	0	0	27	13922	0	-	-
val	BC45	2098	22	22	7	2069	0	4	detected
val	BC49	2996	9	9	10	2977	0	4	detected
val	BC60	8333	0	0	2	8331	0	-	-
val	BC145	5157	4	4	0	5153	0	4	detected
val	BC151	16253	33	31	22	16198	2	4	detected
test	HD2	2952	0	0	1	2951	0	-	-
test	HD21	10092	0	0	5	10087	0	-	-
test	HD22	47715	1	1	6	47708	0	-	-
test	HD24	22651	0	0	6	22645	0	-	-
test	HD25	30248	0	0	19	30229	0	-	-
test	HD26	43178	1	0	4	43173	1	-	-
test	BC5	16547	1	1	3	16543	0	-	-
test	BC21	304053	1	1	13	304039	0	-	-
test	BC25	825957	37	35	9	825911	2	4	detected
test	BC30	5314	0	0	0	5314	0	-	-
test	BC31	22313	58	57	11	22244	1	4	detected
test	BC76	22408	7	4	2	22399	3	6	detected
test	BC77	2798	5	2	1	2792	3	8	detected
test	BC91	21219	0	0	16	21203	0	-	-

test	BC104	51524	0	0	2	51522	0	-	-
test	BC115	48387	2	2	4	48381	0	-	-
test	BC125	13882	0	0	1	13881	0	-	-
test	BC132	11261	1	0	2	11258	1	-	-
test	BC136	9428	0	0	0	9428	0	-	-
test	BC146	26277	14	14	1	26262	0	4	detected
test	BC197	66799	6	6	7	66786	0	6	detected
test	BC199	42726	3	3	1	42722	0	-	-
test	BC219	82240	1	0	5	82234	1	-	-
test	BC220	92856	3	3	8	92845	0	-	-
test	LC4	804	8	8	0	796	0	4	detected
test	LC5	421	8	8	0	413	0	4	detected
test	LC6	3779	10	8	7	3762	2	4	detected
test	LC9	7440	0	0	0	7440	0	-	-
test	LC13	353277	120	120	19	353138	0	4	detected
test	LC14	179617	6	6	3	179608	0	4	detected
test	LC15	147247	0	0	11	147236	0	-	-
test	LC16	144722	1	1	13	144708	0	-	-
test	LC21	82911	1	1	25	82885	0	-	-
test	LC22	17646	11	8	11	17624	3	7	detected
test	LC23	15079	7	3	11	15061	4	16	detected
test	LC24	117340	25	10	23	117292	15	7	detected
test	LC25	12749	12	12	18	12719	0	4	detected
test	LC27	15890	5	3	6	15879	2	10	detected
test	LC28	13628	2	2	3	13623	0	-	-
test	PDAC3	158844	0	0	9	158835	0	-	-
test	PDAC4	289056	6	6	76	288974	0	10	detected
test	PDAC7	463648	6	6	93	463549	0	4	detected
test	PDAC12	334368	0	0	11	334357	0	-	-
test	PDAC3	34279	21	19	95	34163	2	5	detected
test	PDAC4	28726	0	0	6	28720	0	-	-
test	PDAC12	225612	29	26	161	225422	3	5	detected
test	CRC1	2729	0	0	4	2725	0	-	-
test	CRC2	1295	0	0	4	1291	0	-	-
test	CRC3	536515	0	0	7	536508	0	-	-
test	CRC4	9970	1	1	101	9868	0	-	-
test	CRC1	4757	15	5	5	4737	10	5	detected
test	CRC2	1441	24	7	21	1396	17	13	detected
test	CRC3	138751	298	121	136	138317	177	4	detected

test	CRC4	15384	0	0	9	15375	0	-	-
test	RC13	649227	3	3	1304	647920	0	-	-
test	RC22	384	0	0	0	384	0	-	-
test	RC24	120451	95	92	216	120140	3	10	detected
test	RC29	701111	1	1	77	701033	0	-	-
test	RC32	353394	0	0	28	353366	0	-	-
test	RC33	1777	0	0	0	1777	0	-	-
test	RC35	135	0	0	0	135	0	-	-
test	RC36	22129	2	0	2	22125	2	-	-
test	RC37	119	1	0	0	118	1	-	-
test	RC42	1227	0	0	0	1227	0	-	-
test	RC43	266	2	0	0	264	2	-	-
test	RC53	273084	3	2	1288	271793	1	-	-
test	RC57	369573	9	9	87	369477	0	8	detected
test	RC59	207873	2	2	663	207208	0	-	-
test	RC61	100296	2	2	13	100281	0	-	-
test	RC62	300195	0	0	31	300164	0	-	-
test	RC13	301508	133	60	433	300942	73	8	detected
test	RC14	453776	524	378	2658	450594	146	16	detected
test	RC22	629929	564	530	10194	619171	34	8	detected
test	RC25	591621	70	54	280	591271	16	7	detected

Table S3 DAPI+CD45/CD31- gating-narrowing experiment

Summary of an experiment aimed at limiting the false positive ratio for RC samples collected using DAPI+ cells gate by narrowing the analysis to cells within CD45+CD45/CD31- gate. Four samples are studied, all inference results from the same Convnext small model, for each of them three folds are shown: full patient sample – all objects from the sample are analysed as in the remainder of the manuscript; DAPI+CD45/CD31- gated sample – only objects within a gate: positive on DAPI and negative on CD45/CD31 are analysed: normalisation for the model is computed from the gated samples; gated sample – full-normalised – only gated objects are analysed using the model, but they are normalised with the normalisation parameters derived from the full sample;

Patient nr	Fold	Type of blood sample	Objects (n)	CTC (n)	TP (n)	FP (n)	FN (n)	TN (n)
RC53	DAPI+ sample	PB	273084	3	2	1288	1	271793
RC53	DAPI+CD45- gated sample	PB	16695	3	2	4	1	16688
RC53	gated sample - full-normalised	PB	16695	3	2	1239	1	15453
RC57	DAPI+ sample	PB	369573	9	9	87	0	369477
RC57	DAPI+CD45- gated sample	PB	7503	9	8	14	1	7480
RC57	gated sample - full-normalised	PB	7503	9	9	50	0	7444
RC59	DAPI+ sample	PB	207873	2	2	663	0	207208
RC59	DAPI+CD45- gated sample	PB	7902	2	1	35	1	7865
RC59	gated sample - full-normalised	PB	7902	2	2	584	0	7316
RC22	DAPI+ sample	TDVB	629929	564	530	10194	34	619171
RC22	DAPI+CD45- gated sample	TDVB	24035	512	235	624	277	22899
RC22	gated sample - full-normalised	TDVB	24035	512	478	3767	34	19756

Table S4 CTCs thresholds associations with clinical parameters

Associations of CTC positivity categorized according to different cut-offs to lymph node (N) and/or distant organ (M) involvement, and disease progression; F indicates Fisher exact test (otheriwse Chi square test), neg - negativity, pos - positivity

	CTC >=1		CTC >=2		CTC >=3		CTC >=4		CTC >=5	
	neg	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**	**pos**	**neg**	**pos**
no N or M involvement	12 / 63.2%	17 / 30.9 %	15 / 51.7 %	14 / 31.1 %	18 / 51.4%	11 / 28.2%	19 / 50.0 %	10 / 27.8 %	19 / 48.7 %	10 / 28.6 %
N or M involvement	7 / 36.8%	38 / 69.1 %	14 / 48.3 %	31 / 68.9 %	17 / 48.6 %	28 / 71.8%	19 / 50.0 %	26 / 72.2 %	20 / 51.3 %	25 / 71.4 %
	p=0.016 (F)		p=0.076		p=0.041		p=0.050			p=0.076
no disease progression	15 / 68.2 %	28 / 63.8 %	18 / 69.2 %	25 / 62.5 %	21 / 67.7 %	22 / 62.9 %	24 / 70.6 %	19 / 59.4 %	24 / 70.6 %	19 / 59.4 %
disease progression	7 / 31.8 %	16 / 36.4 %	8 / 30.8 %	15 / 37.5 %	10 / 32.3 %	13 / 37.1 %	10 / 29.4 %	13 / 40.6 %	10 / 29.4 %	13 / 40.6 %
	p=0.789 (F)		p=0.575		p=0.678		p=0.339			p=0.339

Figures

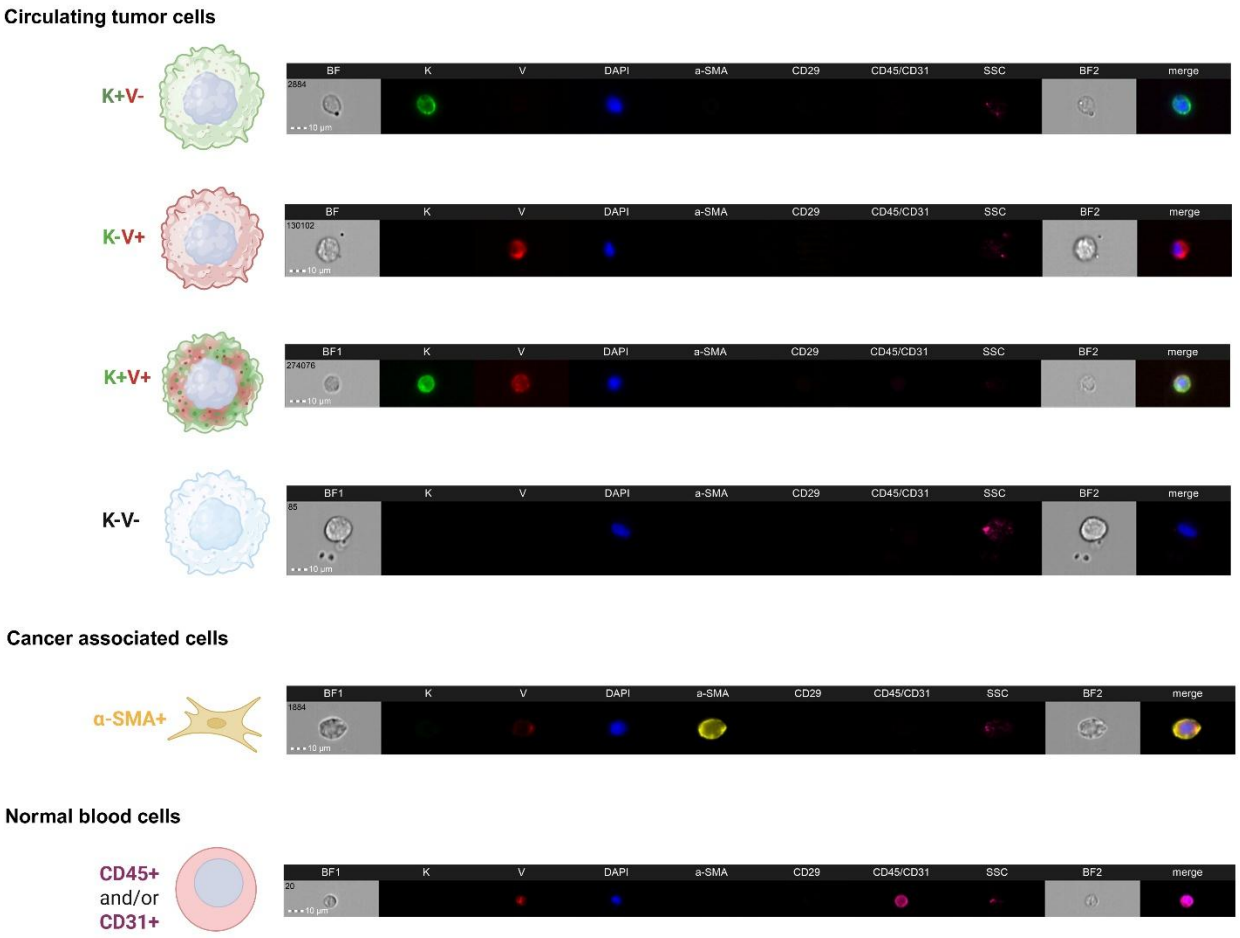


Figure 1 Representative imaging flow cytometry-derived images of circulating tumor cells phenotypes, circulating cancer-associated fibroblasts, and normal blood cells.

Circulating tumor cells are shown according to keratin (K) and vimentin (V) expression as K+V-, K-V+, K+V+, and K-V-. Additional representative cells include an α-smooth muscle actin positive (α-SMA+) cancer-associated fibroblast and a CD45+ and/or CD31+ blood cell. Schematic representations accompany each phenotype. DAPI was used for nuclear staining. BF, brightfield; SSC, side scatter. Created with BioRender.com. Images are acquired using low speed flow and 40x magnification.

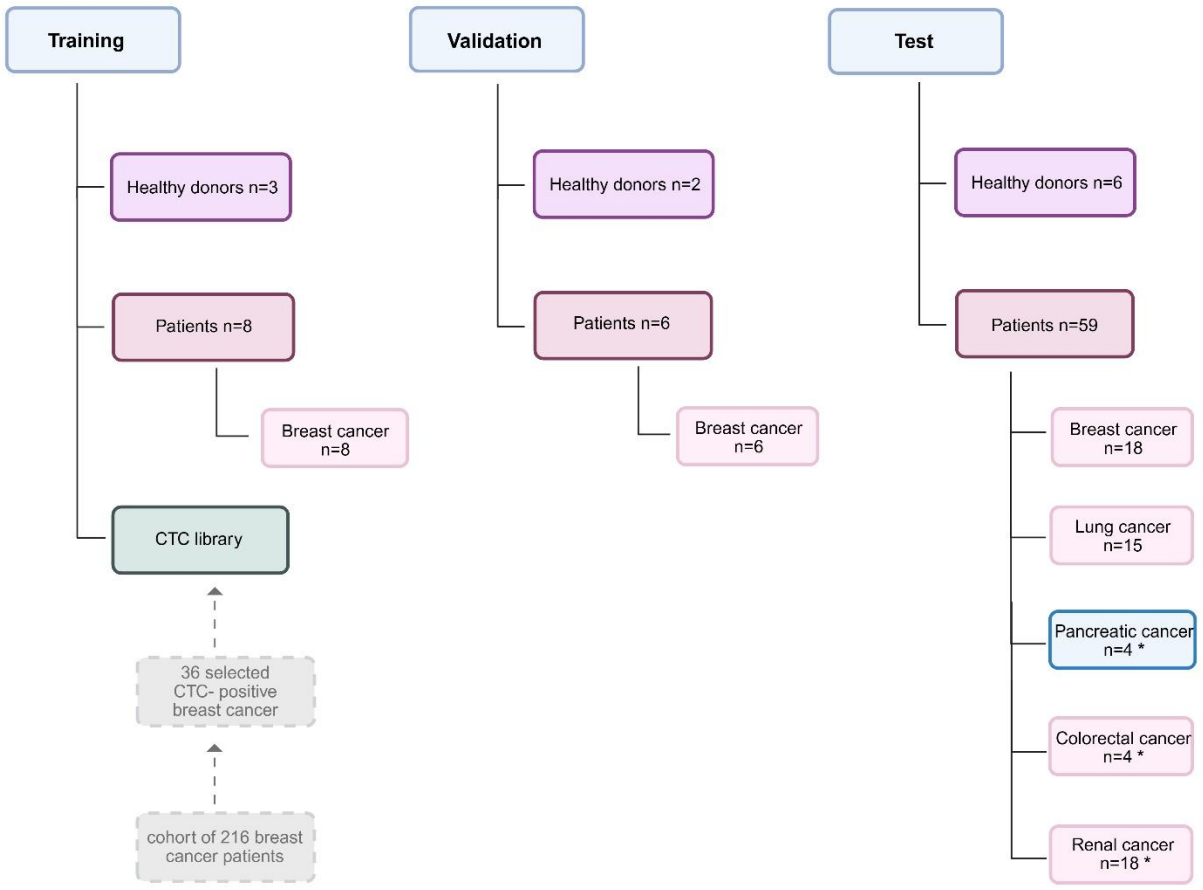


Figure 2 Composition of the training, validation, and test cohorts.

The training cohort included 3 healthy donors and 8 breast cancer patients, supplemented by a circulating tumor cell (CTC) library derived from 36 selected CTC-positive patients from a cohort of 216 breast cancer patients described in detail before. The validation cohort included 2 healthy donors and 6 breast cancer patients. The test cohort included 6 healthy donors and 59 patients with breast (n = 18), lung (n = 15), pancreatic (n = 4), colorectal (n = 4), or renal (n = 18) cancer. Asterisks indicate patient groups in which both peripheral blood (PB) and tumor-draining vein blood (TDVB) were analyzed. The pancreatic cancer group is outlined in blue to distinguish samples collected at the Medical University of Łódź from the remaining samples, which were collected at the University Clinical Center in Gdańsk. Those samples were used to validate the model at an external center using a second instrument. Created with BioRender.com.

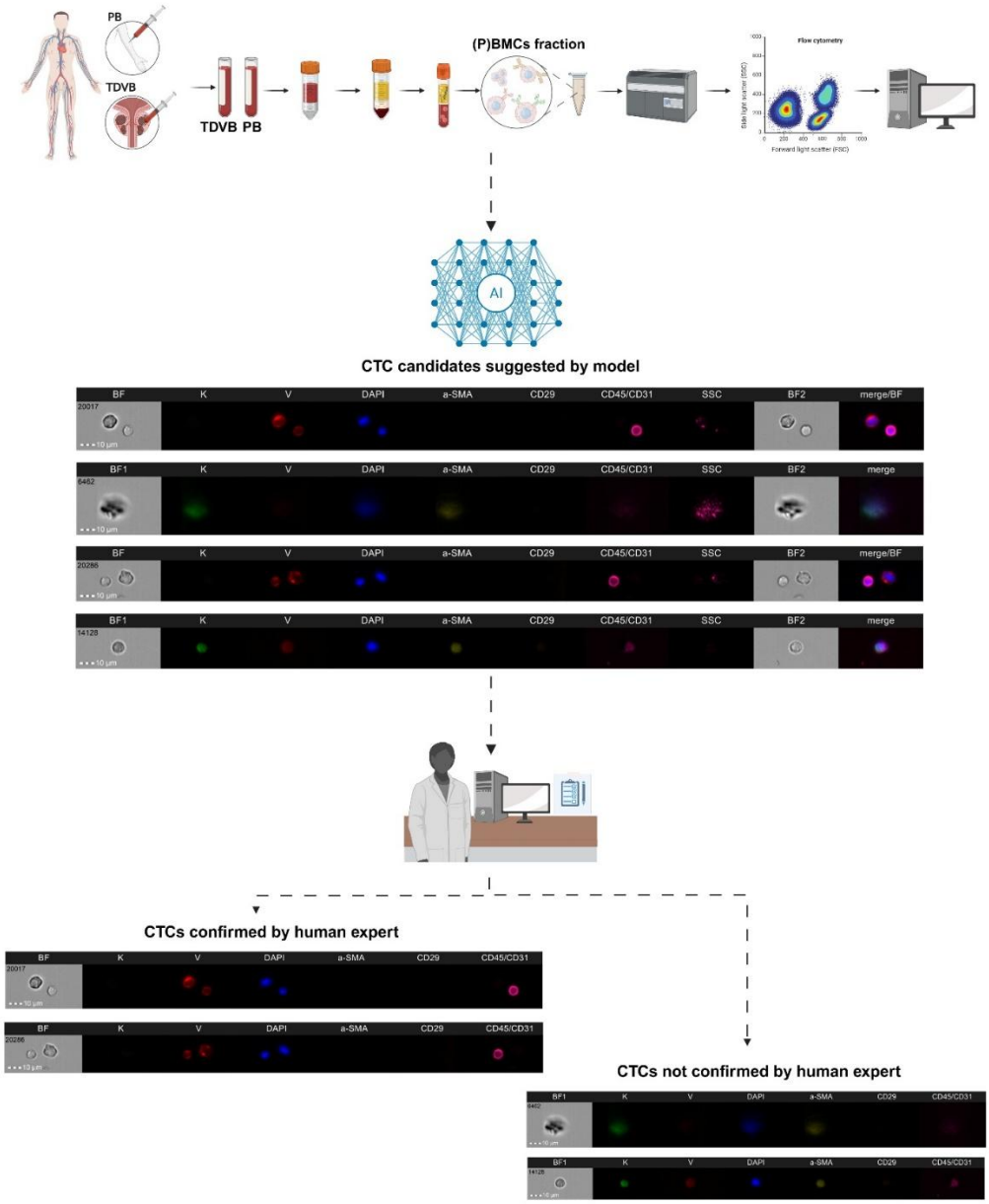


Figure 3 Workflow for artificial intelligence–assisted identification and expert verification of CTCs.

Peripheral blood (PB) and tumor-draining vein blood (TDVB) samples were processed to isolate the (peripheral) mononuclear cell fraction ((P)BMC), followed by immunofluorescence staining and imaging flow cytometry analysis. An artificial intelligence model identified potential CTCs, which were subsequently reviewed by a human expert and classified as confirmed or unconfirmed CTC candidates. Representative images illustrate model-selected candidates and the outcomes of expert review. Created with BioRender.com. The images were acquired using low speed flow and 40x magnification.

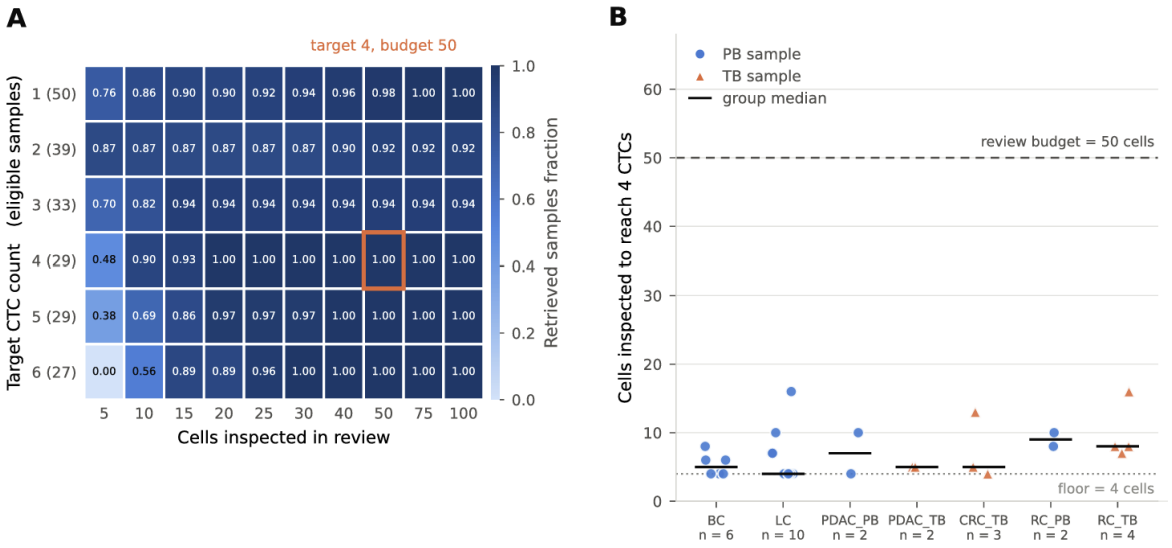


Figure 4 Performance metrics – summary

Performance of the best machine learning model (Convnext Small) is presented. Panel A: Shows the robustness of pairs of thresholds – presenting the ratio of recovered eligible CTCs and CAFs at a given threshold and for a given review budget. The discussed pair of thresholds of 4 and 50, respectively, is highlighted. For a comparison of other models on this metric, see Figure S3. Panel B: Shows for eligible test donor samples (having at least 4 CTCs and CAFs), how many objects need to be reviewed from top model predictions to recover at least 4 CTCs and CAFs; every point is a patient sample, they are grouped per cancer type and blood sample source; the reference review depth of 50 is highlighted. For a comparison of other models on this metric, see Figure S2.

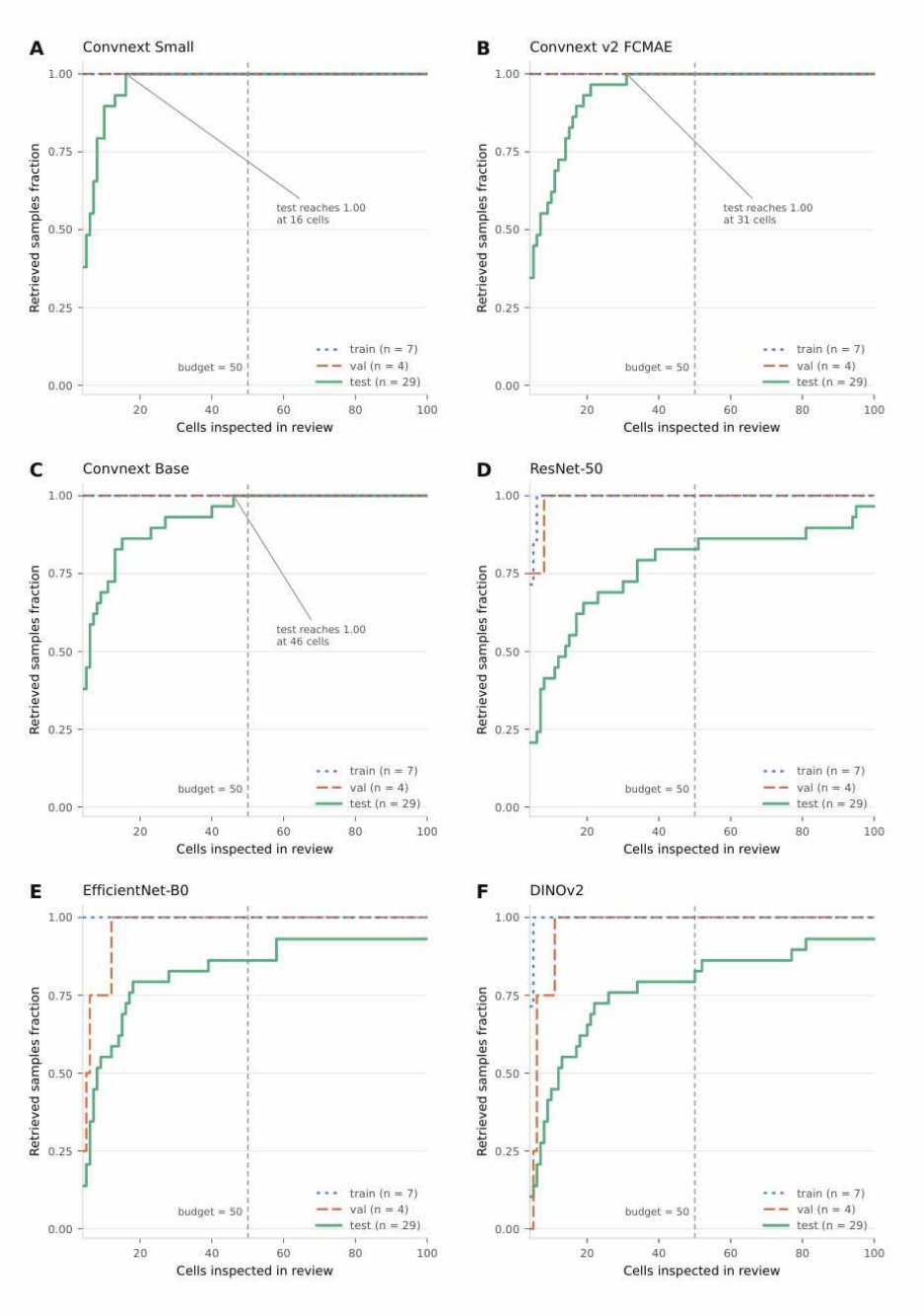
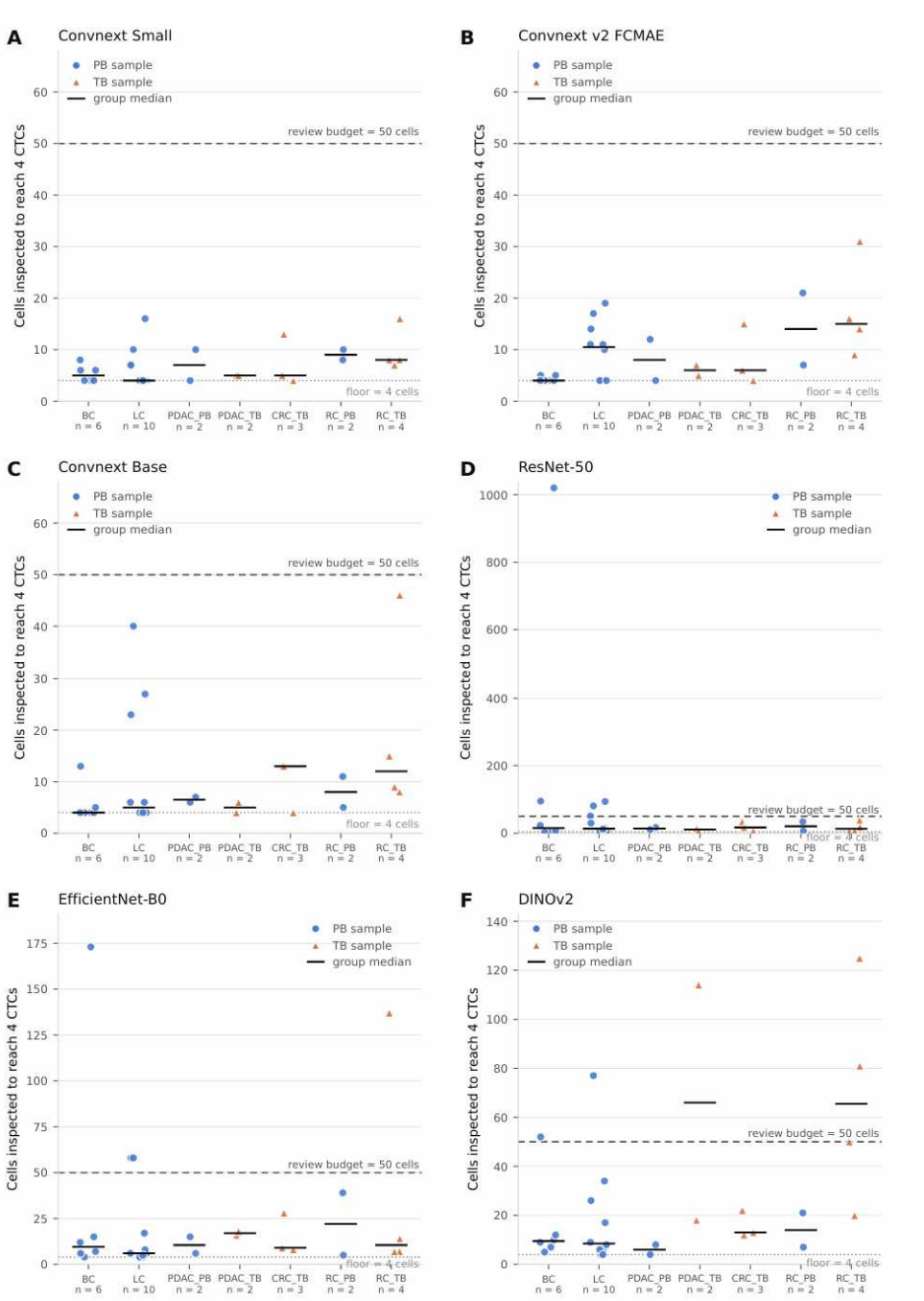


Figure S1 CTC recovery per split – model ablation

Per analysed model, a line is plotted per split (train, val, test) to illustrate at what review depth what combined portion of patient samples is detected, meaning if the sample has at least 4 CTCs and CAFs, at least 4 are within the given review budget. The results illustrate that while all models are able to learn the training set, most struggle to generalise to validation and even more so on test sets. Notably, all compared models from the Convnext family (panels A-C), recover all eligible samples at threshold 4 and



50.

Figure S2 Review depth per sample – model ablation

Per analysed model, eligible test donor samples (having at least 4 CTCs and CAFs), are presented and how many objects need to be reviewed from top model predictions to recover at least 4 CTCs and CAFs; every point is a patient sample, they are grouped per cancer type and blood sample source; the reference review depth of 50 is highlighted.

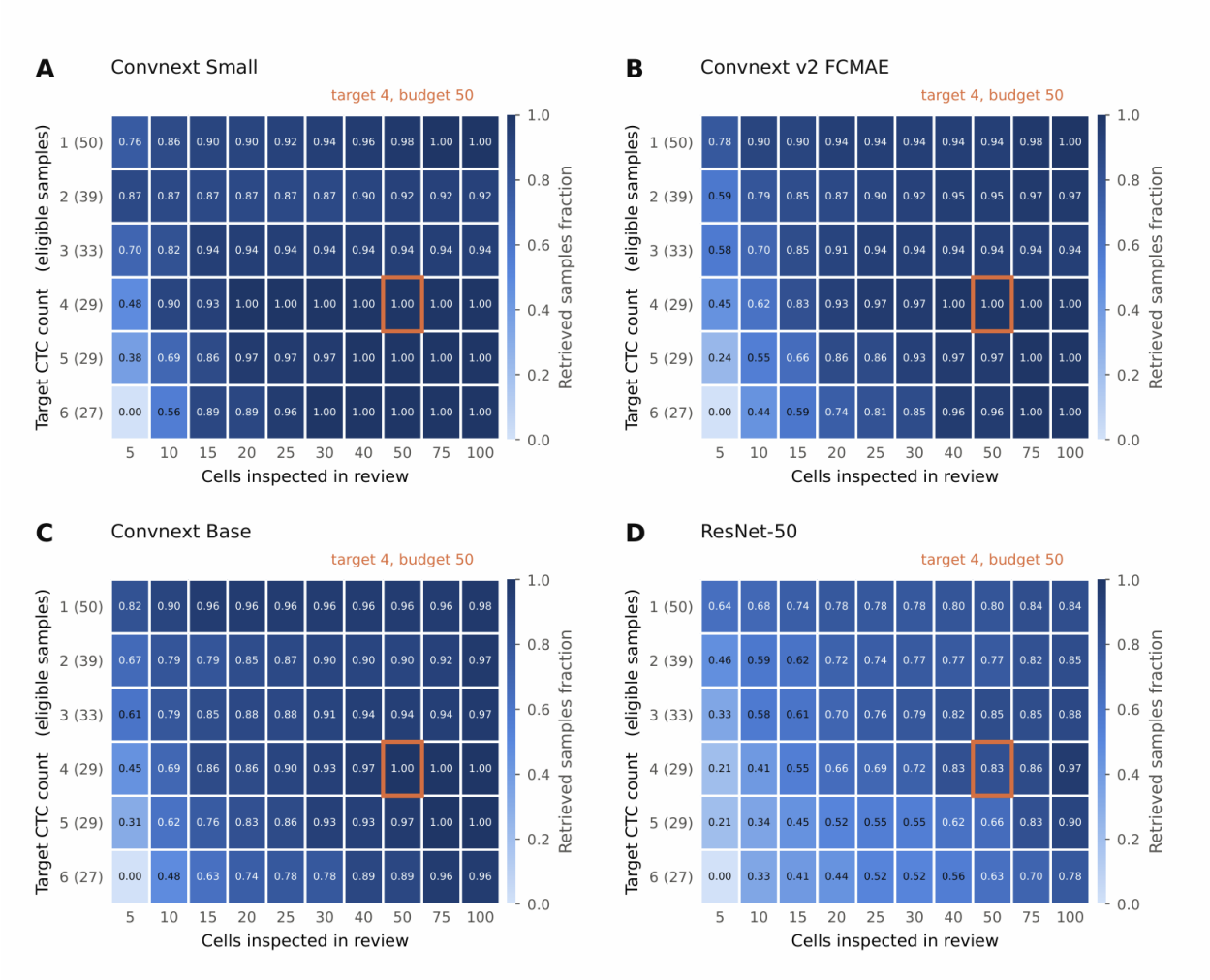


Figure S3 Threshold robustness – model ablation

Per analysed model, a robustness grid of the thresholds is shown, by presenting the ratios of recovered eligible test donor samples at a given threshold and for a given review budget, meaning for samples with at least the threshold number of CTCs and CAFs, at least that number of them is within the given review budget. The selected pair of thresholds of 4 and 50, respectively, is highlighted. Convnext Small is clearly the most robust model from the range.

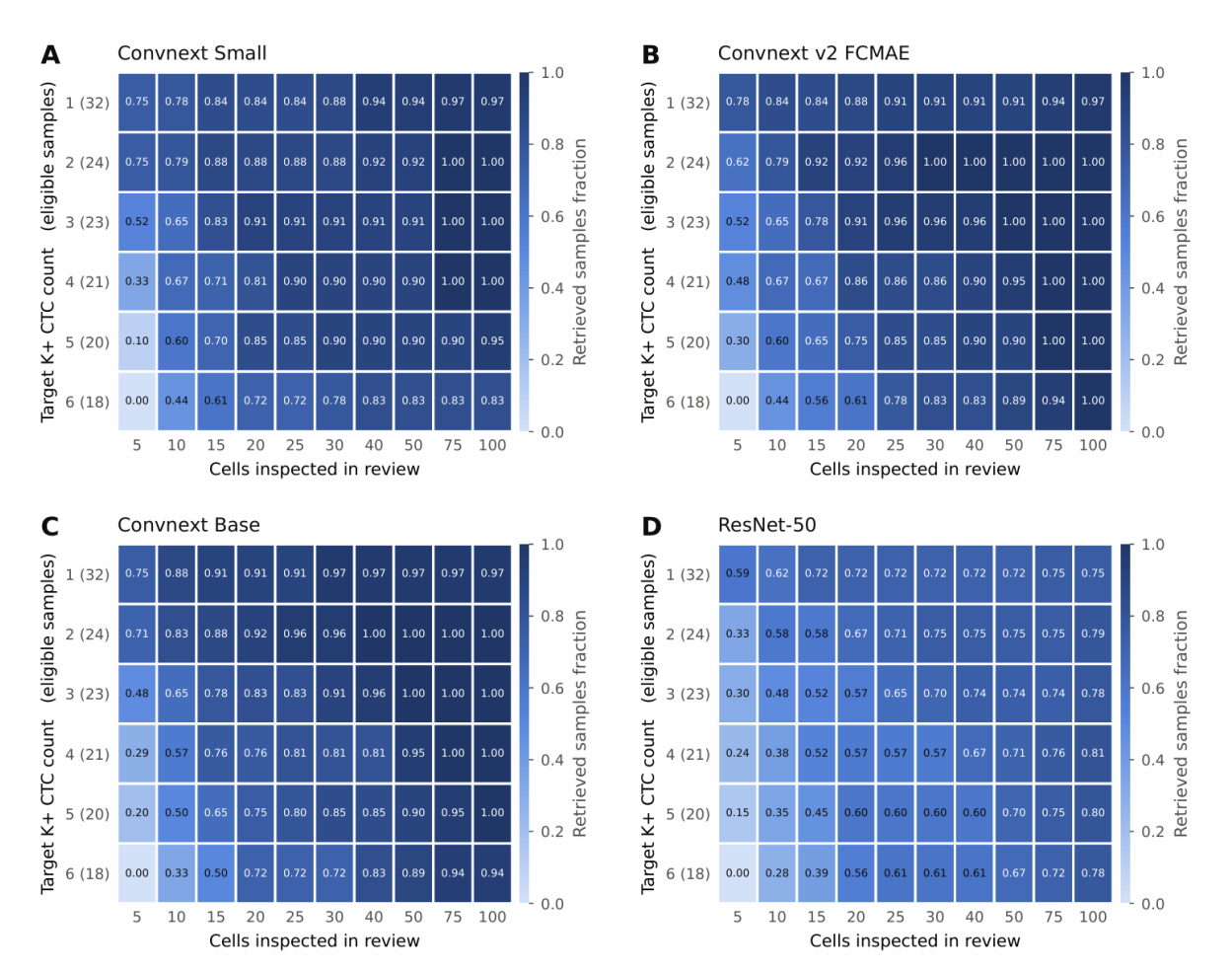


Figure S4 Epithelial-only threshold robustness – model ablation

Similarly, to Figure S3, the threshold robustness is analysed, but taking into account only epithelial CTCs (K+V+ and K+V- populations) all other CTCs and CAFs are treated as negative objects. Using the models trained on identifying all 5 populations for filtering only the epithelial CTCs proves to also be possible, with a more generous review budget of 100 cells. Notably, on this task, Convnext v2 FCMAE (panel B) outperforms Convnext Small (panel A).

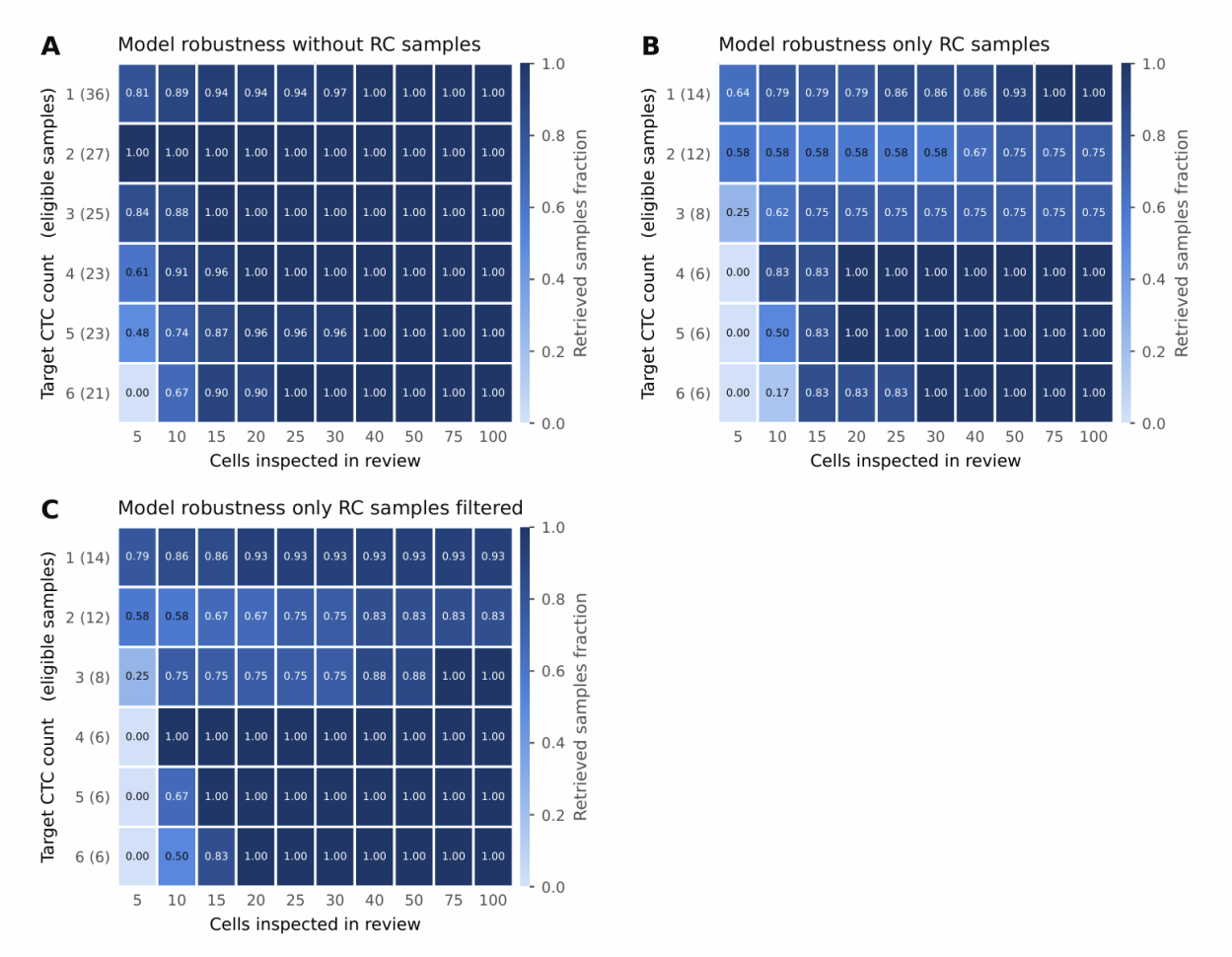


Figure S5 Renal cancer filtering – impact on threshold robustness

The robustness analysis of Convnext Small model presented in Figure 4 panel A, and Figure S3 panel A is split between RC only samples (panel B) and all other test samples (BC, LC, PDAC, CRC, both PB and TDVB) (panel A). It is clear that lack of full robustness for review budgets in the range 40-100 and small thresholds (1-3) is entirely caused by the RC samples. Panel C presents an application of the algorithmic filter on objects with the DAPI and Vimentin channels overlapping in their foreground. Clearly, such an approach can remedy the robustness over the RC samples to some extent.

References

1. Alix-Panabières, C. & Pantel, K. From Discovery to Diagnosis: A Perspective for Circulating Tumor Cells in Personalized Oncology. *Cancer Discov.* **15**, 1985–2001 (2025).
2. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584 (2013).
3. Jordan, N. V. *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102–106 (2016).
4. Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31**, 539–544 (2013).
5. Muchlińska, A. *et al.* Improved Characterization of Circulating Tumor Cells and Cancer-Associated Fibroblasts in One-Tube Assay in Breast Cancer Patients Using Imaging Flow Cytometry. *Cancers* **15**, 4169 (2023).
6. Lin, D. *et al.* Circulating tumor cells: biology and clinical significance. *Signal Transduct. Target. Ther.* **6**, 404 (2021).
7. Bednarz-Knoll, N., Alix-Panabières, C. & Pantel, K. Plasticity of disseminating cancer cells in patients with epithelial malignancies. *Cancer Metastasis Rev.* **31**, 673–687 (2012).
8. Muchlińska, A., Smentoch, J., Żaczek, A. J. & Bednarz-Knoll, N. Detection and Characterization of Circulating Tumor Cells Using Imaging Flow Cytometry-A Perspective Study. *Cancers* **14**, 4178 (2022).
9. Jiang, H. *et al.* Deep learning for computational cytology: A survey. *Med. Image Anal.* **84**, 102691 (2023).
10. Tadavarthi, Y. *et al.* Overview of Noninterpretive Artificial Intelligence Models for Safety, Quality, Workflow, and Education Applications in Radiology Practice. *Radiol. Artif. Intell.* **4**, e210114 (2022).

11. Shmatko, A., Ghaffari Laleh, N., Gerstung, M. & Kather, J. N. Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. *Nat. Cancer* **3**, 1026–1038 (2022).

12. Zeune, L. L. *et al.* Deep learning of circulating tumour cells. *Nat. Mach. Intell.* **2**, 124–133 (2020).

13. Wang, S. *et al.* Label-free detection of rare circulating tumor cells by image analysis and machine learning. *Sci. Rep.* **10**, 12226 (2020).

14. Guo, Z. *et al.* Circulating Tumor Cell Identification Based on Deep Learning. *Front. Oncol.* **12**, 843879 (2022).

15. Zhou, M. *et al.* Precisely Enumerating Circulating Tumor Cells Utilizing a Multi-Functional Microfluidic Chip and Unique Image Interpretation Algorithm. *Theranostics* **7**, 4710–4721 (2017).

16. Bootsma, M. L. *et al.* Longitudinal Phenotyping of Circulating Tumor Cells Using a Scalable Deep Learning Framework. *Clin. Cancer Res.* **32**, 4988–4999 (2026).

17. Wenta, R. *et al.* Measurable morphological features of single circulating tumor cells in selected solid tumors-A pilot study. *Cytom. Part J. Int. Soc. Anal. Cytol.* **105**, 883–892 (2024).

18. Wenta, R. *et al.* Size Matters: Small Circulating Tumor Cells Indicate Worse Therapy Response in Breast Cancer. *MedComm* **6**, e70468 (2025).

19. Rack, B. *et al.* Circulating Tumor Cells Predict Survival in Early Average-to-High Risk Breast Cancer Patients. *JNCI J. Natl. Cancer Inst.* **106**, (2014).

20. De Bono, J. S. *et al.* Circulating Tumor Cells Predict Survival Benefit from Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clin. Cancer Res.* **14**, 6302–6309 (2008).

21. Cristofanilli, M. *et al.* Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).

22. Cohen, S. J. *et al.* Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients With Metastatic Colorectal Cancer. *J. Clin. Oncol.* **26**, 3213–3221 (2008).
23. Ligthart, S. T. *et al.* Circulating Tumor Cells Count and Morphological Features in Breast, Colorectal and Prostate Cancer. *PLoS ONE* **8**, e67148 (2013).
24. Nanou, A. *et al.* Training an automated circulating tumor cell classifier when the true classification is uncertain. *PNAS Nexus* **3**, pgae048 (2024).
25. Husseini-Wüsthoff, H. *et al.* Cluster-based human-in-the-loop strategy for improving machine learning-based circulating tumor cell detection in liquid biopsy. *Patterns* **6**, 101285 (2025).
26. Mallery, K. *et al.* Circulating tumor cell detection in cancer patients using in-flow deep learning holography. *Npj Biosensing* **3**, 23 (2026).
27. Shen, C. *et al.* Automatic detection of circulating tumor cells and cancer associated fibroblasts using deep learning. *Sci. Rep.* **13**, 5708 (2023).
28. Liu, Z. *et al.* A ConvNet for the 2020s. in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 11966–11976 (IEEE, New Orleans, LA, USA, 2022). doi:10.1109/CVPR52688.2022.01167.
29. Woo, S. *et al.* ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 16133–16142 (IEEE, Vancouver, BC, Canada, 2023). doi:10.1109/CVPR52729.2023.01548.
30. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (IEEE, Las Vegas, NV, USA, 2016). doi:10.1109/CVPR.2016.90.

31. Tan, M. & Le, Q. V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.
32. Oquab, M. *et al.* DINOv2: Learning Robust Visual Features without Supervision. Preprint at https://doi.org/10.48550/ARXIV.2304.07193 (2023).