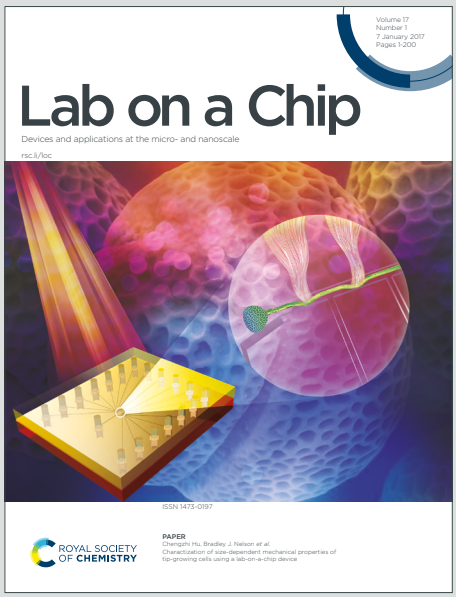
Lab on a Chip

Devices and applications at the micro- and nanoscale

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: M. T. Doganay, P. Chakraborty, S. M. Bommakanti, S. Jammalamadaka, D. Battalapalli, A. Madabhushi and M. Draz, *Lab Chip*, 2024, DOI: 10.1039/D4LC00671B.



ROYAL SOCIETY OF CHEMISTRY

rsc.li/loc

View Article Online
DOI: 10.1039/D4LC00671B

Artificial Intelligence Performance in Testing Microfluidics for Point-of-Care

Mert Tunca Doganay[1], Purbali Chakraborty[1], Sri Moukthika Bommakanti[1], Soujanya Jammalamadaka[1], Dheerendranath Battalapalli[1], Anant Madabhushi[2,3], Mohamed S. Draz[1,4,5*]

1. Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA.
2. Department of Biomedical Engineering, Emory University, Atlanta, GA, USA.
3. Atlanta Veterans Administration Medical Center, Atlanta, GA, USA.
4. Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA.
5. Department of Biomedical Engineering, Cleveland Clinic, Cleveland, OH, 44106, USA.

Corresponding author: Mohamed S. Draz. Email: mohamed.draz@case.edu

View Article Online
DOI: 10.1039/D4LC00671B

ABSTRACT

Artificial intelligence (AI) is revolutionizing medicine by automating tasks like image segmentation and pattern recognition. These AI approaches support seamless integration with existing platforms, enhancing diagnostics, treatment, and patient care. While recent advancements have demonstrated AI superiority in advancing microfluidics for point of care (POC) diagnostics, a gap remains in comparative evaluations of AI algorithms in testing microfluidics. We conducted a comparative evaluation of AI models specifically for the two-class classification problem of identifying the presence or absence of bubbles in microfluidic channels under various imaging conditions. Using a model microfluidic system with a single channel loaded with 3D transparent objects (bubbles), we challenged each of the tested machine learning (ML) (n = 6) and deep learning (DL) (n = 9) models across different background settings. Evaluation revealed that the Random Forest ML model achieved 95.52% sensitivity, 82.57% specificity, and 97% AUC, outperforming other ML algorithms. Among DL models suitable for mobile integration, DenseNet169 demonstrated superior performance, achieving 92.63% sensitivity, 92.22% specificity, and 92% AUC. Remarkably, DenseNet169 integration into a mobile POC system demonstrated exceptional accuracy (> 0.84) in testing microfluidics at under challenging imaging settings. Our study confirms the transformative potential of AI in healthcare, emphasizing its capacity to revolutionize precision medicine through accurate and accessible diagnostics. The integration of AI into healthcare systems holds promise for enhancing patient outcomes and streamlining healthcare delivery.

View Article Online
DOI: 10.1039/D4LC00671B

INTRODUCTION

The convergence of artificial intelligence (AI) and healthcare has opened up a new era of possibilities, particularly in detection diagnostics and treatment. With AI algorithms continuously advancing, the integration of these approaches into healthcare systems holds immense promise for transforming traditional practices and addressing longstanding challenges in healthcare delivery.[1-3] Healthcare applications driven by sophisticated machine learning (ML) and deep learning (DL) algorithms stand at the forefront of modern healthcare innovation.[4-6] These algorithms empower machines to obtain insights from vast datasets, predict clinical outcomes, and assist healthcare providers in making informed decisions.[6] From medical imaging analysis to personalized treatment strategies, AI-driven approaches have demonstrated significant efficacy in improving diagnostic precision and ultimately enhancing patient outcomes.[7-10]

POC diagnostics represent a cornerstone of modern healthcare, offering timely and accessible testing solutions, particularly in resource-limited settings.[11-13] The integration of AI into microfluidic systems presents a promising avenue for enhancing the accessibility and efficiency of POC testing.[14, 15] By harnessing advanced ML and DL algorithms, AI enhances the sensitivity, specificity, and multiplexing capabilities of microfluidic devices, enabling rapid and accurate detection of a wide range of diseases and biomarkers directly at the POC.[16-18] An important approach where AI is utilized to enhance microfluidic systems is in image processing. ML and DL learning models excel at image classification and pattern recognition tasks and can support microfluidic devices to perform rapid and multiplex assays, allowing for comprehensive screening or testing using minimal resources.[19-21] This integration addresses critical gaps in healthcare access and empowers a new level of POC diagnostics, equipping frontline providers with actionable insights and revolutionizing the delivery of healthcare services.

View Article Online
DOI: 10.1039/D4LC00671B

Recent advancements have demonstrated superior performance in identifying disease biomarkers, detecting cancer,[22] viruses,[23] bacteria,[24] and other pathogens,[25] underscoring the robustness and clinical relevance of AI-integrated microfluidic platforms in modern healthcare settings. However, despite these advancements, there remains a gap in the comparative evaluations of different AI algorithms in testing microfluidics, and the optimal approach for maximizing their performance in this context remains unclear, particularly in the POC diagnostics.[26-31] In POC settings, practical constraints such as cost, power consumption, memory limitations, and computational efficiency are crucial, making the choice of algorithm highly impactful. For instance, logistic regression is relatively simple, with a complexity of O(n×m), where *n* is the number of samples and *m* the number of features. It requires moderate computational power and memory, making it a good fit for POC settings that have limited Central Processing Unit (CPU) power and memory.[32] Decision trees, with complexity O(n×m×log(n)),[33] and random forests, which add an additional factor for the number of trees (O(k×n×m×log(n)),[34] where *k* is the number of trees), require moderate resources. They build tree structures that evaluate multiple features at once. While computationally more demanding than logistic regression, they can still be feasible in many POC setups, especially with fewer trees. Naive Bayes classifiers are computationally efficient due to their independence assumption for features, with complexity O(n×m). This makes them ideal in resource-limited environments. However, this simplification can sometimes reduce predictive performance if feature independence is not a valid assumption.[35] On the other hand, Support Vector Machines (SVMs), especially with non-linear kernels, can have significantly higher complexities ($O(n^2)$ to $O(n^3)$), making them less suitable for constrained environments without powerful CPUs or Graphics Processing Units (GPUs). However, using linear kernels or approximation methods (e.g., LinearSVM or Fast SVM) can reduce the computational load,

View Article Online
DOI: 10.1039/D4LC00671B

making SVMs a more viable option for POC.[36] K-Nearest Neighbors (K-NN), while simple in terms of training complexity (O(n×m)), can become computationally intensive during inference due to distance calculations between all data points. Optimization techniques like KD-trees (K-Dimensional trees) or Ball-trees can speed up inference, making K-NN more feasible for real-time POC applications.[37] Neural networks and deep learning models (e.g., Convolutional Neural Networks (CNNs)) typically have a higher complexity of O(n×m×d), where *d* is the depth of the network. These models require substantial memory and processing power, particularly using GPU/TPU resources (where TPU stands for Tensor Processing Units), which are not commonly available in POC devices. However, methods like dropout, batch normalization, weight pruning, and model distillation can help reduce the computational burden, allowing for more lightweight versions of these models to be deployed on smaller devices.[38] Foundation models, like large-scale AI models (e.g., Generative Pre-trained Transformers (GPT), Bidirectional Encoder Representations from Transformers (BERT)), present an even bigger challenge due to their high computational demands during both training and inference. These models often require substantial GPU clusters or high-performance computing (HPC) environments, making them impractical for resource-constrained POC settings. In such cases, pre-trained models fine-tuned for specific tasks or more compact versions of these models (e.g., TinyBERT, DistilBERT) might be used instead.[39] This trade-off between computational demands and resource availability emphasizes the importance of balancing model performance with resource constraints in POC settings.

We employed a model microfluidic system, featuring a single microfluidic channel loaded with 3D transparent objects of bubbles. This model is designed to rigorously challenge the performance of commonly used AI models and provide insights into their effectiveness in real-world diagnostic scenarios. We integrated various ML and DL algorithms into our study, including

View Article Online
DOI: 10.1039/D4LC00671B

CNNs like MobileNetV2, ResNet101V2, and DenseNet169, alongside commonly used ML models in healthcare applications such as Naive Bayes, logistic regression, KNN, SVM, and Random Forest.[40-44] Among the six evaluated ML algorithms, the Random Forest model performed best, achieving 95.52% sensitivity, 82.57% specificity, and 97% AUC. Similarly, among the nine DL models, DenseNet169 stood out, achieving 92.63% sensitivity, 92.22% specificity, and 92% AUC. Such a comparative study is critical in gaining a comprehensive understanding of the strengths and weaknesses of different algorithms, informing algorithm selection, optimization, and deployment decisions across diverse domains and applications.[45-48]

RESULTS AND DISCUSSION

The integration of AI in medicine is driven by its remarkable ability to analyze and classify images and datasets. This computational capability of AI algorithms is foundational across diverse domains, prominently within diagnostics and medical testing, where AI-driven image analysis stands as a transformative force, providing rapid data processing and precise assessment devoid of infrastructure constraints or specialized human oversight.[3, 49, 50] This technological paradigm bears profound implications, particularly on POC diagnostics, through its role in facilitating the integration of microfluidics into POC applications.[51] By harnessing sophisticated ML and DL algorithms, AI streamlines the imaging and analysis of microfluidic devices, such as smartphone-captured assays, reducing the total testing cost and time, enhancing accuracy, and expanding utility.[19, 52, 53] This convergence of AI and microfluidics within POC holds immense potential to

View Article Online
DOI: 10.1039/D4LC00671B

democratize healthcare access, particularly in underserved regions, by providing affordable, accurate, and accessible diagnostic solutions.[14, 19, 54, 55]

In our study, we investigated the efficacy of AI algorithms, including both ML and DL, to facilitate the process of testing microfluidics within POC settings. We employed a microfluidic system comprising a single microfluidic channel to rigorously assess a set of 15 AI models recognized for data analysis and image classification across biomedical and diagnostic domains. Our experimental setup incorporated testing configurations featuring varying densities of bubbles. Bubbles as a readout was selected to probe the imaging and analytical performance of the examined algorithms. Despite bubbles being less prevalent than conventional color-based or fluorescence-based readouts, their inherent 3D transparency poses challenges, as they may be mistaken for non-targeted constituents within the sample matrix, microfluidic system or the testing environment and background. In addition, transparent bubbles can introduce challenges such as refraction and variable light scattering, which may impact imaging accuracy and algorithm performance. By using these bubbles, we aimed to simulate complex real-world imaging conditions and evaluate how well the AI models could handle such complexities. Colorimetric readouts, though linear and would allow comparatively easier workflow, fail to sufficiently encapsulate the intricacies necessary for discerning strengths and weaknesses of the tested algorithms. Meanwhile, fluorescence, although know to support high specificity and sensitivity testing, remains impractical for widespread POC adoption due to the need for bulky equipment and specialized setup to achieve the required sensitivity and specificity in most analyses.

Our set of AI algorithms included ML models, such as Naive Bayes, logistic regression, k-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest, alongside DL CNNs such as MobileNetV2, ResNet101V2, and DenseNet169. By combining traditional ML

View Article Online
DOI: 10.1039/D4LC00671B

algorithms with state-of-the-art CNN architectures, we created a diverse ensemble of models that can collectively leverage different aspects of the data. This ensemble approach is essential to enhance robustness and generalization performance, particularly in scenarios where the dataset may be limited or the target features are challenging to discern (i.e., bubbles). The incorporation of traditional ML algorithms stemmed from their robustness in handling various types of features, including those extracted from images, and their suitability for the often constrained datasets characteristic of microfluidic diagnostics at POC settings. The CNN architectures like MobileNetV2, ResNet101V2, and DenseNet169 have unparalleled ability to capture intricate spatial relationships within images, which is crucial for discerning subtle patterns like challenging signals such as bubbles. This aligns with the evolving field of diagnostics, which is moving towards inventing and incorporating more versatile readouts like bubbles to allow for more sensitive and unique detection capabilities, distinct from common ones like color and fluorescence. These CNN architectures offer distinct trade-offs in terms of model size, computational efficiency, and classification accuracy, offering flexibility in addressing the specific nuances of the dataset.

To investigate the capabilities of the selected set of ML and DL algorithms in testing microfluidics, we captured 19,097 images of our microfluidic model with bubbles in various settings, including different environments, lighting conditions, times of the day, and backgrounds (**Figure 1**). We labeled the captured images either positive or negative, based on the number of bubbles, around a threshold value of 10 bubbles per microchip, to train our ML and DL models (**Figure 1a**). Out of the 19,097 labelled images (**Figure 1b**), 15,530 images were utilized for training using Python running on Lambda Vector GPU Workstation (Intel i9-10900x CPU, NVIDIA RTX A6000 GPU) system.

View Article Online
DOI: 10.1039/D4LC00671B

To test the performance of ML models, we used 1595 randomly selected images, excluding those used for training, to evaluate their classification accuracy. We employed standard performance metrics, including accuracy, precision, recall (i.e., sensitivity), specificity, F1 score, and Matthews's correlation coefficient (MCC) (**Supplementary Table 1**), obtained from each model to determine their effectiveness.[56] We conducted all statistical analyses and data visualizations using TensorFlow and TensorBoard tools with necessary Python libraries as matplotlib, NumPy, Keras, Sklearn, pandas, torch.[57, 58] The comparison primarily centered around specificity and sensitivity values, which are metrics influencing overall performance and gives information about other metrics.

Our analysis of the ML models revealed that logistic regression and Random Forest models exhibited exceptional sensitivity (>90%), while K-nearest neighbors and Random Forest models demonstrated high specificity (>80%) (**Figure 2a**). The results showed that the highest sensitivity value was obtained from the Random Forests (95.52%) and the highest specificity value was obtained from K-nearest neighbors (89.68%) ML models. we assessed the confusion matrix to better understand the positive and negative predictions. Out of 1595 images, 1447 were classified correctly, with 45 false negatives and 103 false positives. The model primarily made errors in the classification of negative samples. (**Figure 2b** and **Supplementary Figure 1**). The ROC analysis of the trained models indicated that the Random Forest (AUC: 97%) (**Figure 2c**) and K-nearest neighbors (AUC: 90%) have highest area under the ROC, which represents the diagnostic ability of the model (**Supplementary Figure 2**). Additionally, the Random Forest model outperformed others in terms of F1 score (92.8%) and accuracy (90.72%). This shows that the Random Forest provides most balanced results between precision and sensitivity with highest accuracy. Consequently, the most effective model was observed as Random Forest with notable metrics as

View Article Online
DOI: 10.1039/D4LC00671B

95.52% sensitivity, 82.57% specificity, 90.72% accuracy, 90.3% precision, 92.8% F1 score, 79.95% MCC, and 97% AUC (**Supplementary Table 1**).

To test the performance of DL models, we continued by evaluating the performance of the selected CNNs architectures using the same dataset of 1595 images. The performance evaluation step was conducted using developed Python algorithms with the help of Pandas, NumPy, Sklearn, Matplotlib, Keras and Tensorflow libraries.[57] The deep learning models utilized for this evaluation included MobileNetV2, EfficientNetV2B0, EfficientNetV2B2, DenseNet169, DenseNet201, InceptionV3, ResNet50V2, EfficientNetB5, and ResNet101V2. In selecting these deep learning models, we prioritized those that does not require significant computing power and thus ensure compatibility for evaluation and testing microfluidics at POC. We also ensured that the chosen models were commonly employed for computer vision tasks, prioritizing ease of integration and robust performance on POC compatible mobile devices.[19]

Our results indicated that DenseNet169, EfficientNetB5, and EfficientNetV2B0 exhibited outstanding sensitivity values of 92.63%, 95.82%, and 91.93%, respectively (**Figure 3a and Supplementary Figure 3-5**). ResNet50V2 (89.17%) and InceptionV3 (88.49%) demonstrated high specificity values, while DenseNet169 displayed an exceptional specificity of 92.22% (**Supplementary Table 2**). The confusion matrix revealed further insights into the performance of these algorithms. DenseNet169 algorithm excelled in detecting negative samples, accurately classifying 545 out of 591, while also achieving the second-highest performance in positive classification with 930 out of 1004, resulting in the highest overall performance at 92% (**Figure 3b**). Other algorithms including EfficientNetB5 correctly identified 962 out of the tested 1004 positive samples. However, it misclassified 293 negative samples as positive, resulting in a 50.4% performance rate for negative samples and an overall performance rate of 79%. EfficientNetV2B0

View Article Online
DOI: 10.1039/D4LC00671B

exhibited similar performance, albeit with a 7% overall performance rate downgrade, reflecting a 4% difference in true positive performance rate and an 11% decrease in true negative performance rate. The results of MobileNetV2, EfficientNetV2B2, DenseNet201, InceptionV3, ResNet50V2, and ResNet101V2 algorithms are shown in Supplementary **Figure 4, 5** with misclassification rates < 38%. The ROC analysis of the trained DL models, ResNet50V2 (AUC: 96%), ResNet101V2 (AUC: 96%), InceptionV3 (AUC: 95%) and DenseNet169 (AUC: 92%) and DenseNet201 (AUC: 90%) had the highest area under the ROC (**Supplementary Figure 6, 7**). Additionally, the DenseNet169 model outperformed other models in terms of F1 score (93.94%) and accuracy (92.48%) (**Supplementary Table 2**). Overall, DenseNet169 outperformed other models with the performance metrics and gives the applicable model with 0.92 AUC (**Figure 3c**).

We compared the performance of Random Forest and DenseNet169, as these models had outperformed others in our evaluations. To challenge them further, we used a set of 184 microchips prepared with varying numbers of bubbles. A new test set of images was created under different environmental conditions than those used during training. This test set included images taken against different backgrounds (including black, red, brown, metallic grey, and dark blue), rotation, and brightness. This approach allowed us to assess user experience in suboptimal conditions, ensuring a thorough and comprehensive evaluation of the models' performance in real-world microchip testing scenarios. The generated positive and negative prediction rates were analyzed against the ground truth values of bubbles per chip to evaluate the performance of each model. The results revealed that the DenseNet169 DL model achieves prediction rates with better performance compared to the Random Forest ML model with 80.4% and 88.2% accuracy; 77.98% and 91.81% precision; 81.51% and 87.84% F1 score; 75.3% and 92.31% specificity; and 61.03% and 76.69% MCC for Random Forest and DenseNet169, respectively. The confusion matrix and ROC analyses,

View Article Online
DOI: 10.1039/D4LC00671B

on the other hand, confirmed that the DenseNet169 DL algorithm is the optimal prediction model for testing our microfluidic model, outperforming the Random Forest ML algorithm by 87% in AUC and 92% in accuracy classifying true positive and true negative (**Figure 4b, c**).

To demonstrate the effectiveness of incorporating AI in real-world sample testing scenarios using POC-compatible systems, a mobile application capable of running the DenseNet169 model seamlessly was developed, without the need for further optimization. The application features a simple interface for initiating model evaluation and presents results in terms of positive and negative prediction rates, along with images of the tested microfluidic chips (**Supplementary Figure 8**). Out of 250 images, 212 were classified correctly, 29 were classified as false negatives, and 9 were classified as false positives. The model primarily made errors in classifying positive samples. The performance metrics were as follows: Accuracy: 84.8%, Precision: 93.23%, Sensitivity/Recall: 81.05%, F1 Score: 86.71%, Specificity: 90.72%, and MCC: 70.09. The deep learning model achieved an AUC value of 0.90, highlighting its superiority in testing our microfluidic model with bubbles (**Figure 5b**). Furthermore, upon examining the confusion matrix alongside sensitivity and specificity values. Results showed that the DenseNet169 deep learning model achieved 81.05% sensitivity and 90.72% specificity (**Figure 5a**). Heatmap analysis was conducted using images with bubble counts ranging from 0 to 100. The results indicated a higher margin of error around the threshold of 10 bubbles, particularly chips with around 20 to 30 bubbles are ~30 % misclassified as negative.

Our study provides a comprehensive evaluation of both ML and deep learning DL algorithms in the context of microfluidics testing under POC settings. Among the ML models, Random Forest emerged as the top performer with a sensitivity of 95.52%, specificity of 82.57%, and an AUC of 97%, showcasing its strong capability in accurately classifying microfluidic device

View Article Online
DOI: 10.1039/D4LC00671B

images. The high sensitivity and specificity values underscore Random Forest's effectiveness in distinguishing positive from negative samples even in challenging imaging conditions. However, the higher rate of false positives indicates a potential area for improvement. In contrast, DL models, particularly DenseNet169, exhibited outstanding performance with sensitivity and specificity values of 92.63% and 92.22%, respectively. DenseNet169's consistent high performance across different testing conditions, including variations in background and lighting, highlights its robustness and adaptability, making it highly suitable for real-world POC diagnostics where consistent and reliable performance is crucial.

Despite the promising results, several challenges must be addressed to facilitate the widespread adoption of AI in microfluidic POC diagnostics. One key issue is the misclassification of samples with a marginal number of bubbles, especially around the threshold of 10 bubbles, which was evident in the heatmap analysis. Further refinement of the AI models and incorporating additional features or training data will be necessary to enhance accuracy in borderline cases. Combining multiple algorithms can also help overcome these challenges. For example, employing ensemble techniques that integrate models like U-Net for image segmentation and Canny edge detection for edge detection could improve precision in detecting subtle features. Additionally, integrating algorithms such as YOLO (You Only Look Once) for real-time object detection and HOG (Histogram of Oriented Gradients) for robust feature extraction can further enhance the accuracy and reliability of microfluidic POC diagnostics. Such hybrid approaches can leverage the strengths of different algorithms, providing a more comprehensive and accurate analysis.

Moreover, integrating AI models into mobile applications for POC testing will necessitate ensuring seamless operation across a wide range of devices and environmental conditions, with a strong emphasis on user-friendliness and reliability. This integration is pivotal for achieving the

View Article Online
DOI: 10.1039/D4LC00671B

robustness required for practical deployment in diverse healthcare settings. The successful implementation of AI in microfluidic POC diagnostics has far-reaching implications for the healthcare industry, especially in resource-limited settings where access to sophisticated medical infrastructure is often constrained. By enabling rapid, accurate, and on-site testing, AI-driven POC systems address one of the most pressing challenges in modern medicine: the need for timely and precise diagnostics. By democratizing access to high-quality diagnostic tools, AI-integrated POC systems empower frontline healthcare providers with actionable insights, fostering a more equitable distribution of medical resources. This shift supports personalized medicine approaches, tailoring treatment plans to individual patient profiles based on accurate and immediate diagnostic data. Ultimately, the widespread adoption of AI-enhanced microfluidic POC diagnostics can transform healthcare delivery, making it more accessible, efficient, and responsive to the needs of diverse populations worldwide.

Conclusion

The transformative impact of AI on healthcare is rapidly increasing, particularly in advancing precision medicine through accurate and accessible diagnostics. By conducting a comprehensive comparative evaluation of AI models in testing microfluidics, we have demonstrated the superiority of AI-driven approaches over traditional methods, particularly in the context of POC diagnostics. Through the integration of ML and DL algorithms, we created a diverse ensemble of models capable of leveraging various aspects of the data, thereby enhancing robustness and generalization performance. Our results revealed that the Random Forest ML model and the DenseNet169 DL model exhibited exceptional performance, surpassing other

View Article Online
DOI: 10.1039/D4LC00671B

algorithms in terms of sensitivity, specificity, and AUC values. DenseNet169 integration into a mobile POC system demonstrated exceptional accuracy, outperforming traditional visual interpretation by a significant margin. This confirms the potential of AI to revolutionize diagnostics, offering more accurate and efficient testing solutions in resource-limited settings. Moreover, our findings highlight the significant role that AI can play into healthcare systems, as it holds promise for enhancing patient outcomes, streamlining healthcare delivery, and ultimately, democratizing access to high-quality diagnostic services. Moving forward, further research and development efforts are warranted to optimize AI algorithms for real-world deployment, ensuring their seamless integration into clinical practice and maximizing their impact on global health outcomes.

Data availability

The data supporting this article have been included as part of the ESI.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors would like to thank Dr. Cyril R. A. John Chelliah, Ebenesh Chandrakumar, Dheeksha Devaraj, and Aravinth Kumar for useful discussions and running initial trials.

View Article Online
DOI: 10.1039/D4LC00671B

MATERIAL AND METHODS

Microfluidic chip model design and fabrication.

We developed a microfluid chip system that features a single microfluidic channel. The microchip was designed using the vector graphics editor CorelDRAW Graphics suite software, and fabricated of polymethyl methacrylate (PMMA) (3.125 mm thick), DSA film (100 μm thick, 3M, USA), and glass slides (25 mm x 75 mm). The fabrication process starts by cutting PMMA and DSA film using a laser cutter (Boss Laser LS-1416, USA). The PMMA was prepared to contain the microfluidic channel inlet and outlet, while DSA film included the main testing channel. All materials were precleaned with 70% ethanol, and deionized water using lint-free tissue. The surface of the cleaned glass slides was treated and cleaned using oxygen plasma (PE-25, 100 mW, 15% oxygen; Plasma Etch Inc.) for 10 minutes. Then PMMA and DSA film were assembled on the modified glass slide, forming the model microfluidic chip system. Each system was loaded with platinum nanoparticle-seeded bubbles. PtNPs synthesized using our previously published protocol were mixed with a peroxide-containing solution (5% hydrogen peroxide and 20% glycerol) and loaded on chip system. The concentration of added PtNPs was controlled to prepare systems with variable numbers of bubbles (0 – >200 bubbles per chip), randomly distributed within the microfluidic channel.

AI models selection, training and performance testing

We selected a set of 15 models that encompass a number of machine learning and deep learning models, widely reported to have high performance in image classification and pattern recognition. The machine learning models included Naive Bayes, Logistic Regression, Decision Tree, K-Nearest Neighbors, Support Vector Machine and Random Forest, while the deep learning

View Article Online
DOI: 10.1039/D4LC00671B

models of MobileNetV2, EfficientNetV2B0, EfficientNetV2B2, DenseNet169, DenseNet201, InceptionV3, ResNet50V2, EfficientNetB5 and ResNet101V2, were selected to support workflow running on mobile devices and systems. We generated a dataset of 19,097 images of the model microfluidic system captured using Moto XT1575, iPhone X and Vivo smartphones. The dataset comprises two groups, i.e., positive (with > 10 bubbles per microchip) and negative (in range of < 10 bubbles per microchip) sample images. The microfluidic system imaging was performed at different angles (0 – 360º) and backgrounds and environments to maximize the variations, and make our dataset more robust and comprehensive. We used 15530 images for training, 1788 images for validation and 1012 images for testing the performance of the selected ML and DL models in testing the model microfluidic system and classifying samples into positive and negative based on bubble signal. We started the process by importing pre-trained models available from Scikit-learn and Keras libraries to develop the selected ML and DL models, respectively. In the pre-processing step, the images of our training dataset were resized to the input dimensions of the selected models, leveraging the features learned by ImageNet pretrained network. We performed the batch normalization then used Adam optimizer to fine-tune the network using a global learning rate of 0.001. In addition, we employed a varied number of epochs to test the algorithms optimal performance and we set the number to 50 epochs. Then we performed the transfer learning by removing the final classification layer from the chosen networks and trained with our dataset. All the algorithms were trained on Vector Workstation (Intel i9-10900x CPU and NVIDIA RTX A6000 GPU, Lambda) and after training, we tested the performance of the best-performing ML and DL algorithms individually using a challenging dataset of 400 images. This testing dataset included rotated images, images with various colored backgrounds (matte, bright, reflective), and images with lens distortion and brightness variations. The ML algorithms were evaluated using

View Article Online
DOI: 10.1039/D4LC00671B

the sklearn and torch libraries, while the DL algorithm was evaluated using the TensorFlow library. Performance metrics such as accuracy, precision, sensitivity, and F1-score were employed to quantitatively measure classification accuracy and the ability of each model to correctly identify the tested microchip.

AI testing on a POC compatible system.

We utilized the open-source platform Android Studio (version Giraffe 2022.3.1) to develop an AI-enabled mobile application. Android Studio offers an integrated development environment (IDE) tailored for Android application development. The application facilitates the capture of sensor images through the smartphone's built-in camera or from images stored in the device's memory. A trained DL model, DenseNet169, was converted to TensorFlow Lite and integrated into the application, which was developed for Android 6.0 (API level 23). This application was installed on a Moto XT1575 and used as a proof-of-concept system for testing microfluidics with images simulating real-world conditions. We evaluated the performance of the AI model using a testing set of 250 images, each featuring 0-100 bubbles per chip. This testing set included images with challenging backgrounds and imaging conditions, such as noise, blur, hand interaction, daylight, artificial light, natural and artificial occlusion, resolution variability, and the presence of small bubbles. The classification results, displayed on the user interface, indicate the probability of a sample being positive (>50%) or negative (<50%). The correlation between AI-generated classification results and the number of bubbles per chip was analyzed, and prediction accuracy rates were employed to generate performance metrics.

View Article Online
DOI: 10.1039/D4LC00671B

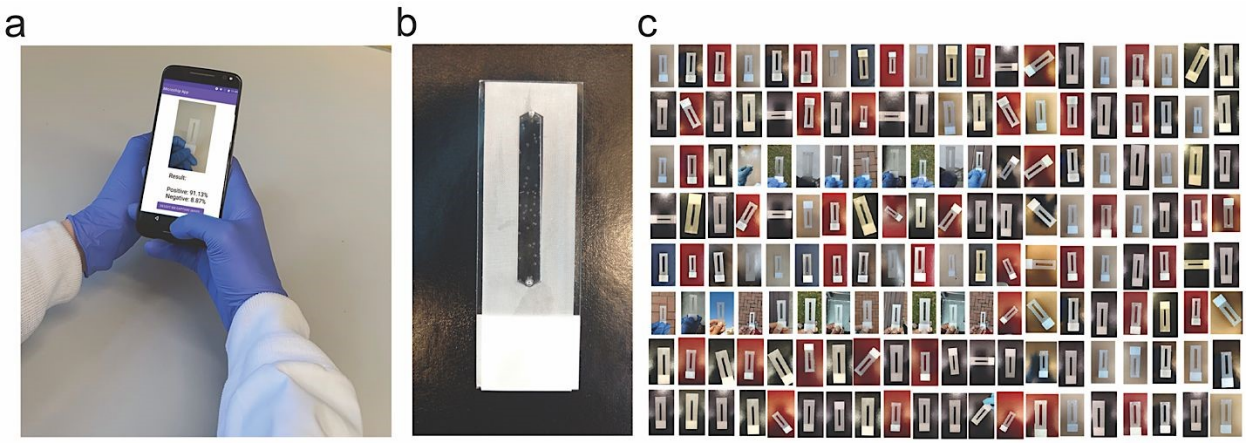


Figure 1. **AI algorithms integration and the tested microfluidic model system.** (**a**) Microfluidics testing using an integrated POC compatible system running AI algorithm on a cellphone. The system supports a broad range of AI algorithms including both machine learning (ML) and deep learning (DL) models. (**b**) The developed microfluidic model with a single microfluidic channel (length 42 mm, width 5 mm and height 100 μm) containing platinum nanoparticle-seeded bubbles of variable shapes and sizes. (**c**) Snapshot of the image library of the tested microfluidic model collected using cellphone POC system (161 randomly selected images out of 19,097), illustrating the diversity of color, background and brightness.

View Article Online
DOI: 10.1039/D4LC00671B

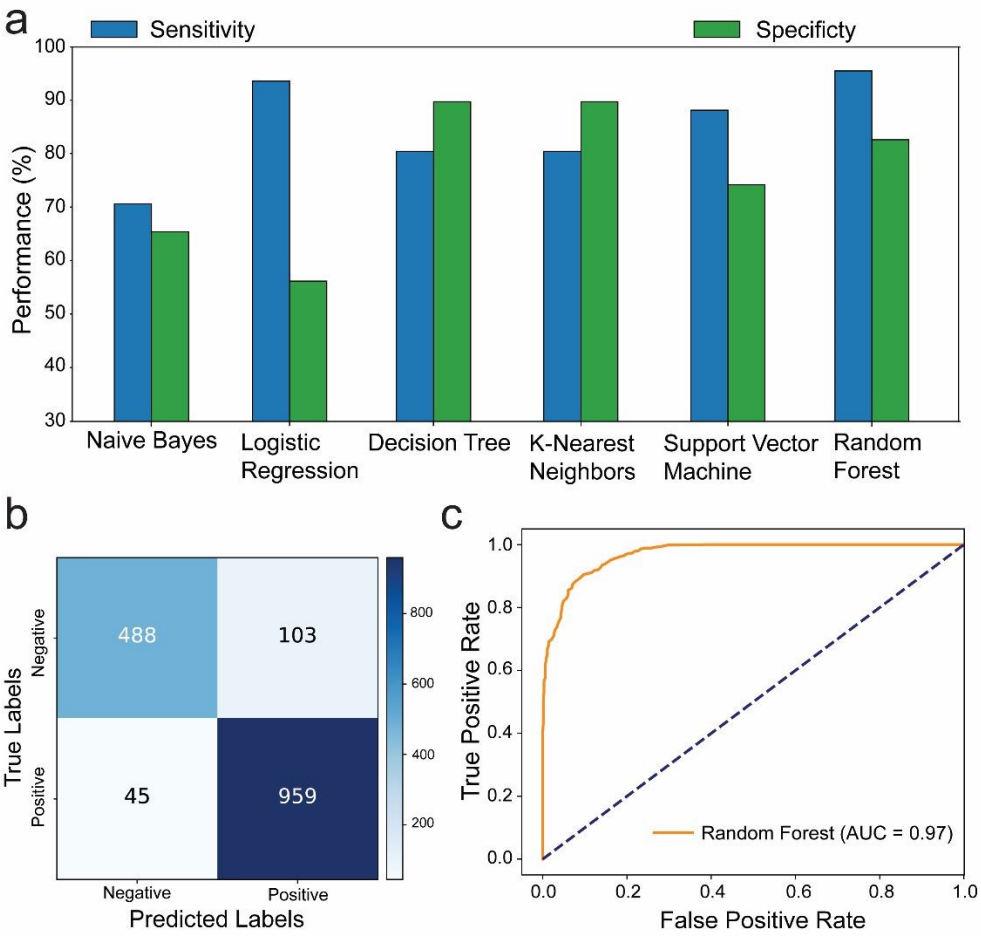


Figure 2. Performance evaluation of machine learning in testing microfluidics. (**a**) Barplots showing the performance (sensitivity and specificity) of the tested ML algorithms (n = 6). All algorithms were trained on our dataset of 15,530 images to classify the model microfluidic chip system with bubble signal into positive or negative around the threshold value of 10 bubbles. (**b**) Confusion matrix showing the number of true negative, false positive, false negative and true positive results when comparing the interpretation of Random Forest ML algorithm to the ground truth classification results. (**c**) ROC analysis of Random Forest performance in testing the model microfluidic chip with bubble signal.

View Article Online
DOI: 10.1039/D4LC00671B

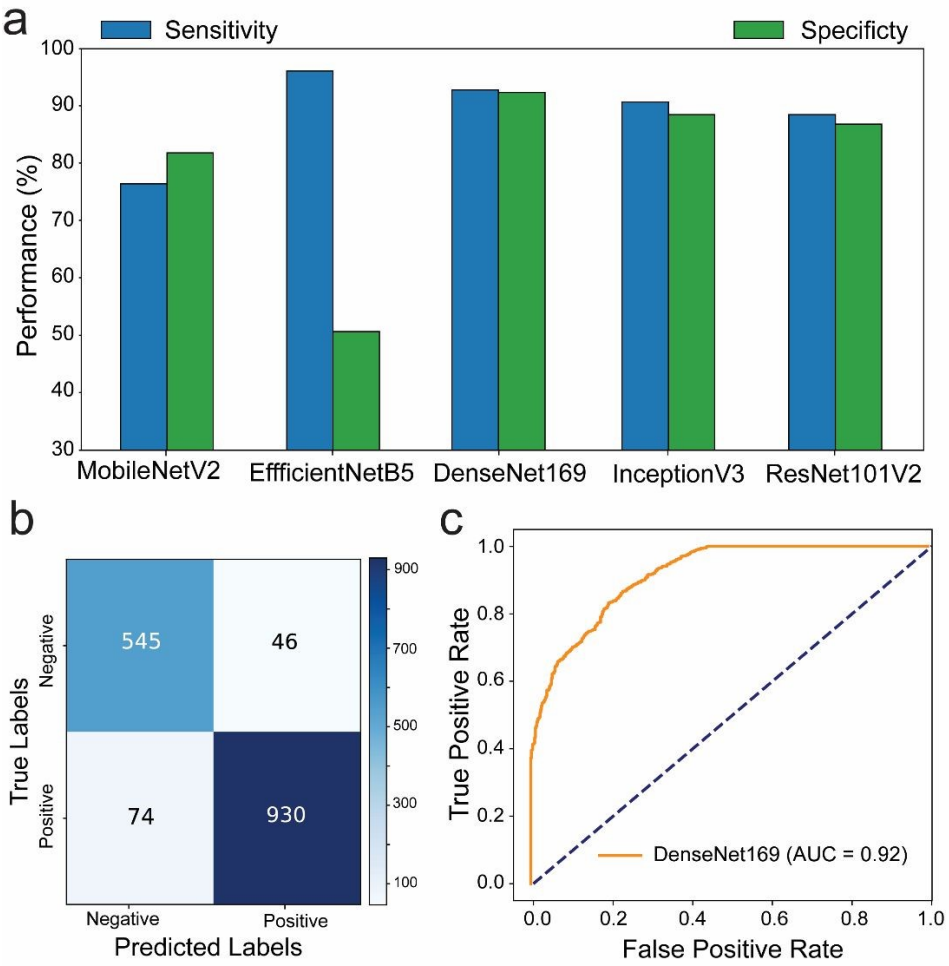


Figure 3. **Performance evaluation of deep learning in testing microfluidics.** (**a**) Barplots showing the performance (sensitivity and specificity) of the tested DL algorithms (n = 5). All algorithms were trained on our dataset of 15,530 images to classify the model microfluidic chip system with bubble signal into positive or negative around the threshold value of 10 bubbles. (**b**) Confusion matrix showing the number of true negative, false positive, false negative and true positive results when comparing the interpretation of DenseNet169 DL algorithm to the ground truth classification results. (**c**) ROC analysis of DenseNet169 performance in testing the model microfluidic chip system with bubble signal.

View Article Online
DOI: 10.1039/D4LC00671B

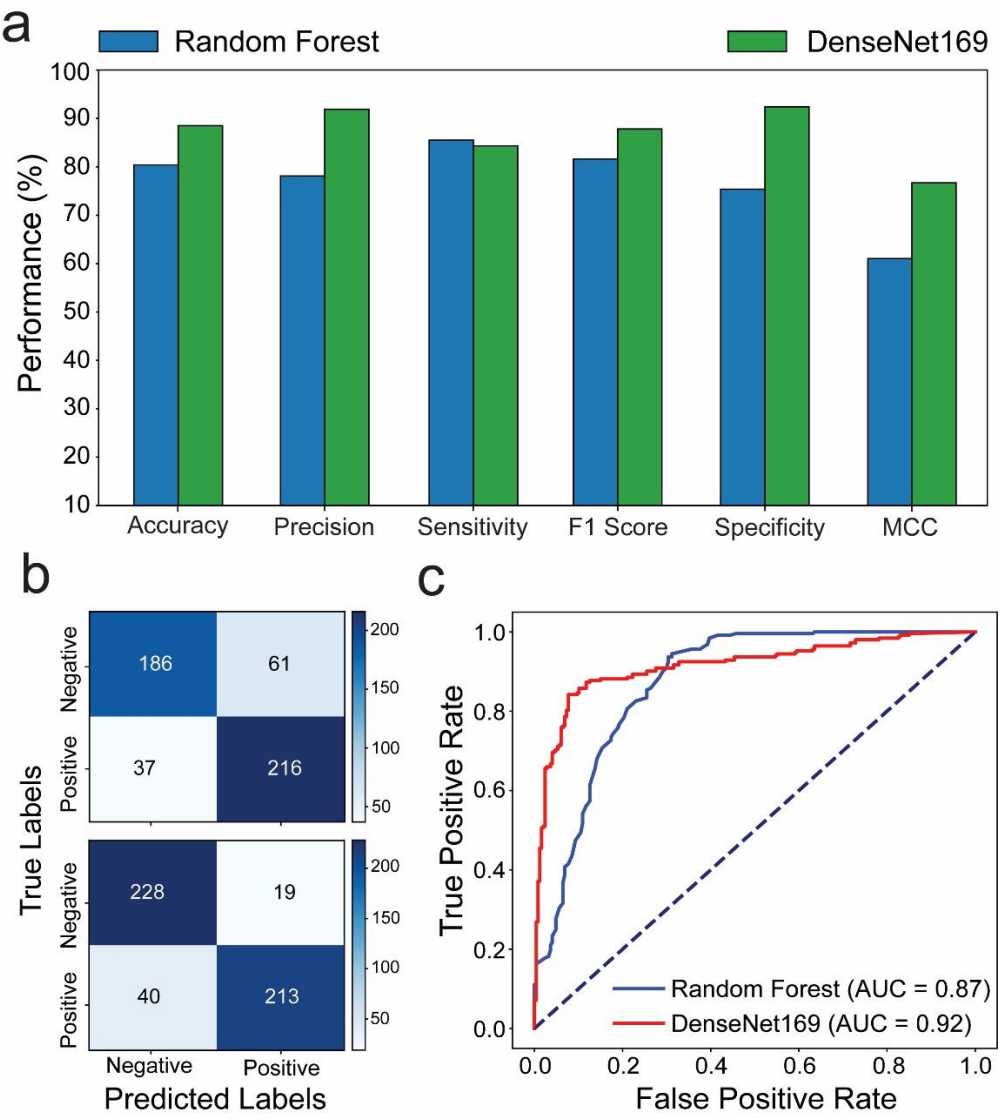


Figure 4. Performance evaluation of machine learning compared to deep learning in testing microfluidics under POC settings. (a) Performance matrices (accuracy, precision, sensitivity, F1 score, specificity, and MCC) of the Random Forest ML and the DenseNet169 DL in testing the model microfluidic chip system under challenging imaging conditions that simulate POC testing settings (i.e., different backgrounds, brightness, resolution, cameras, and rotations). (**b**) Confusion matrices showing the number of true negative, false positive, false negative and true positive results when comparing the interpretation of the Random Forest ML and the DenseNet169 DL algorithms to the ground truth classification results. (**c**) ROC analysis of the Random Forest ML and the DenseNet169 DL algorithms performance in testing the model microfluidic chip system with bubble signal.

View Article Online
DOI: 10.1039/D4LC00671B

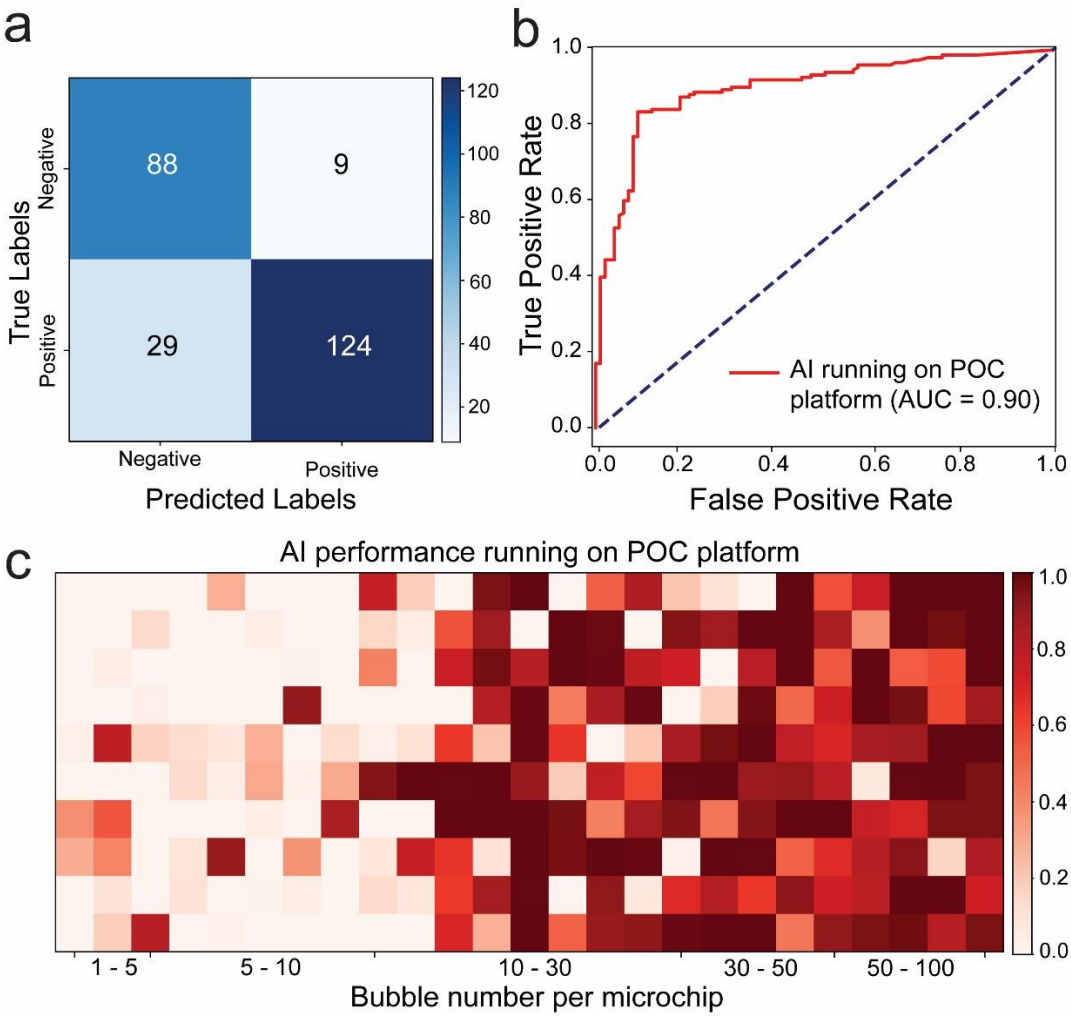


Figure 5. Performance evaluation of AI in testing microfluidics under POC settings using a compatible cellphone system. (**a**) The confusion matrix showing the number of true negative, false positive, false negative and true positive results when comparing AI (i.e., the DenseNet169 DL algorithm) interpretation to the ground truth classification results based on the number of bubbles per microchip. (**b**) ROC analysis of AI performance in testing the model microfluidic chip system with bubble signal. (**c**) Heatmap plot of the probability values of the model microfluidic testing interpretation by AI performance based on the number of bubbles per microchip.

View Article Online
DOI: 10.1039/D4LC00671B

References

1. P. Rajpurkar, E. Chen, O. Banerjee and E. J. Topol, *Nature Medicine*, 2022, **28**, 31-38.
2. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun and J. Dean, *Nature medicine*, 2019, **25**, 24-29.
3. E. J. Topol, *Nature Medicine*, 2019, **25**, 44-56.
4. J. N. Acosta, G. J. Falcone, P. Rajpurkar and E. J. Topol, *Nature Medicine*, 2022, **28**, 1773-1784.
5. A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz and H. J. W. L. Aerts, *Nature Reviews Cancer*, 2018, **18**, 500-510.
6. D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu and F. Yan, *cell*, 2018, **172**, 1122-1131. e1129.
7. R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian and A. Darzi, *npj Digital Medicine*, 2021, **4**, 65.
8. P. Lambin, R. T. H. Leijenaar, T. M. Deist, J. Peerlings, E. E. C. de Jong, J. van Timmeren, S. Sanduleanu, R. T. H. M. Larue, A. J. G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger and S. Walsh, *Nature Reviews Clinical Oncology*, 2017, **14**, 749-762.
9. M. Sermesant, H. Delingette, H. Cochet, P. Jaïs and N. Ayache, *Nature Reviews Cardiology*, 2021, **18**, 600-609.
10. O. Oren, B. J. Gersh and D. L. Bhatt, *The Lancet Digital Health*, 2020, **2**, e486-e488.
11. P. Yager, G. J. Domingo and J. Gerdes, *Annu. Rev. Biomed. Eng.*, 2008, **10**, 107-144.
12. C. P. Y. Chan, W. C. Mak, K. Y. Cheung, K. K. Sin, C. M. Yu, T. H. Rainer and R. Renneberg, *Annual review of analytical chemistry*, 2013, **6**, 191-211.
13. C. Wang, M. Liu, Z. Wang, S. Li, Y. Deng and N. He, *Nano Today*, 2021, **37**, 101092.
14. J. Riordon, D. Sovilj, S. Sanner, D. Sinton and E. W. Young, *Trends in biotechnology*, 2019, **37**, 310-324.
15. S. Chen, Z. Qiao, Y. Niu, J. C. Yeo, Y. Liu, J. Qi, S. Fan, X. Liu, J. Y. Lee and C. T. Lim, *Nature Reviews Bioengineering*, 2023, **1**, 950-971.
16. J. Zhou, J. Dong, H. Hou, L. Huang and J. Li, *Lab on a Chip*, 2024.
17. W. Zhao, Y. Zhou, Y.-Z. Feng, X. Niu, Y. Zhao, J. Zhao, Y. Dong, M. Tan, Y. Xianyu and Y. Chen, *ACS nano*, 2023, **17**, 13700-13714.
18. Z. Ao, H. Cai, Z. Wu, L. Hu, A. Nunez, Z. Zhou, H. Liu, M. Bondesson, X. Lu and X. Lu, *Proceedings of the National Academy of Sciences*, 2022, **119**, e2214569119.
19. B. Wang, Y. Li, M. Zhou, Y. Han, M. Zhang, Z. Gao, Z. Liu, P. Chen, W. Du and X. Zhang, *Nature Communications*, 2023, **14**, 1-18.
20. H. Liu, L. Nan, F. Chen, Y. Zhao and Y. Zhao, *Lab on a Chip*, 2023, **23**, 2497-2513.
21. J. Zheng, T. Cole, Y. Zhang, J. Kim and S.-Y. Tang, *Biosensors and Bioelectronics*, 2021, **194**, 113666.
22. W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout, C. Abbosh and I. F. Dunn, *CA: a cancer journal for clinicians*, 2019, **69**, 127-157.
23. N. Arora, A. K. Banerjee and M. L. Narasu, *Journal*, 2020, **15**, 717-724.
24. K. P. Smith, H. Wang, T. J. Durant, B. A. Mathison, S. E. Sharp, J. E. Kirby, S. W. Long and D. D. Rhoads, *Clinical Microbiology Newsletter*, 2020, **42**, 61-70.
25. K. P. Smith and J. E. Kirby, *Clinical Microbiology and Infection*, 2020, **26**, 1318-1323.
26. A. J. London, *Hastings Center Report*, 2019, **49**, 15-21.
27. C. Gilvary, N. Madhukar, J. Elkhader and O. Elemento, *Trends in pharmacological sciences*, 2019, **40**, 555-564.

View Article Online
DOI: 10.1039/D4LC00671B

28. J. E. Dayhoff and J. M. DeLeo, *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 2001, **91**, 1615-1635.
29. O. Koteluk, A. Wartecki, S. Mazurek, I. Kołodziejczak and A. Mackiewicz, *Journal of Personalized Medicine*, 2021, **11**, 32.
30. H. Bhaskar, D. C. Hoyle and S. Singh, *Computers in biology and medicine*, 2006, **36**, 1104-1125.
31. A. A. de Hond, A. M. Leeuwenberg, L. Hooft, I. M. Kant, S. W. Nijman, H. J. van Os, J. J. Aardoom, T. P. Debray, E. Schuit and M. van Smeden, *NPJ digital medicine*, 2022, **5**, 2.
32. M. Sumner, E. Frank and M. Hall, 2005.
33. H. M. Sani, C. Lei and D. Neagu, 2018.
34. X. Zheng, J. Jia, S. Guo, J. Chen, L. Sun, Y. Xiong and W. Xu, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, **14**, 2222-2235.
35. Z. Zheng, 1998.
36. L. Bottou and C.-J. Lin, 2007.
37. K. Hajebi, Y. Abbasi-Yadkori, H. Shahbazi and H. Zhang, 2011.
38. Y. Zhang, L. Wang, J. Zhao, X. Han, H. Wu, M. Li and M. Deveci, *Information Sciences*, 2024, **670**, 120644.
39. R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut and E. Brunskill, *arXiv preprint arXiv:2108.07258*, 2021.
40. R. C. Deo, *Circulation*, 2015, **132**, 1920-1930.
41. A. Alanazi, *Informatics in Medicine Unlocked*, 2022, **30**, 100924.
42. M. A. Morid, A. Borjali and G. Del Fiol, *Computers in biology and medicine*, 2021, **128**, 104115.
43. A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui and A. Mellit, *Sustainability*, 2023, **15**, 5930.
44. M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu and X. Zhao, *Journal of X-ray Science and Technology*, 2020, **28**, 821-839.
45. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, *Proceedings of the IEEE*, 2020, **109**, 43-76.
46. S. J. Pan and Q. Yang, *IEEE Transactions on knowledge and data engineering*, 2009, **22**, 1345-1359.
47. J. Ooge, G. Stiglic and K. Verbert, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2022, **12**, e1427.
48. K. Moulaei, A. Yadegari, M. Baharestani, S. Farzanbakhsh, B. Sabet and M. R. Afrash, *International Journal of Medical Informatics*, 2024, 105474.
49. X. Tang, *BJR| Open*, 2019, **2**, 20190031.
50. N. Hasani, M. A. Morris, A. Rahmim, R. M. Summers, E. Jones, E. Siegel and B. Saboury, *PET clinics*, 2022, **17**, 1-12.
51. Z. Angehrn, L. Haldna, A. S. Zandvliet, E. Gil Berglund, J. Zeeuw, B. Amzal, S. A. Cheung, T. M. Polasek, M. Pfister and T. Kerbusch, *Frontiers in pharmacology*, 2020, **11**, 759.
52. Y. Yang, F. Xu, J. Chen, C. Tao, Y. Li, Q. Chen, S. Tang, H. K. Lee and W. Shen, *Biosensors and Bioelectronics*, 2023, 115233.
53. D. Xu, X. Huang, J. Guo and X. Ma, *Biosensors and Bioelectronics*, 2018, **110**, 78-88.
54. D. McIntyre, A. Lashkaripour, P. Fordyce and D. Densmore, *Lab on a Chip*, 2022, **22**, 2925-2937.
55. I. Hernández-Neuta, F. Neumann, J. Brightmeyer, T. Ba Tis, N. Madaboosi, Q. Wei, A. Ozcan and M. Nilsson, *Journal of internal medicine*, 2019, **285**, 19-39.
56. Y. Jiao and P. Du, *Quantitative Biology*, 2016, **4**, 320-330.
57. I. Stančin and A. Jović, 2019.
58. B. J. Erickson, P. Korfiatis, Z. Akkus, T. Kline and K. Philbrick, *Journal of digital imaging*, 2017, **30**, 400-405.

View Article Online
DOI: 10.1039/D4LC00671B

547

View Article Online
DOI: 10.1039/D4LC00671B

Artificial Intelligence Performance in Testing Microfluidics for Point-of-Care

Mert Tunca Doganay[1], Purbali Chakraborty[1], Sri Moukthika[1], Soujanya Jammalamadaka[1], Dheerendranath Battalapalli[1], Anant Madabhushi[2,3], Mohamed S. Draz[1,4,5*]

1. Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA.
2. Department of Biomedical Engineering, Emory University, Atlanta, GA, USA.
3. Atlanta Veterans Administration Medical Center, Atlanta, GA, USA.
4. Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA.
5. Department of Biomedical Engineering, Cleveland Clinic, Cleveland, OH, 44106, USA.

Corresponding author: Mohamed S. Draz. Email: mohamed.draz@case.edu

View Article Online
DOI: 10.1039/D4LC00671B

Data Availability

The authors confirm that all data supporting the findings of this study are included within the article and its supplementary information (ESI).