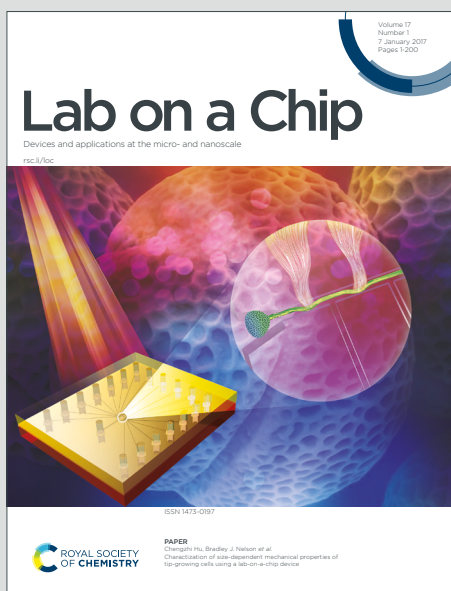


Lab on a Chip

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.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

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

2 Mert Tunca Doganay¹, Purbali Chakraborty¹, Sri Moukthika Bommakanti¹, Soujanya

3 Jammalamadaka¹, Dheerendranath Battalapalli¹, Anant Madabhushi^{2,3}, Mohamed S. Draz^{1,4,5*}

4 1. Department of Medicine, Case Western Reserve University School of Medicine, Cleveland,
5 OH, 44106, USA.

6 2. Department of Biomedical Engineering, Emory University, Atlanta, GA, USA.

7 3. Atlanta Veterans Administration Medical Center, Atlanta, GA, USA.

8 4. Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH,
9 USA.

10 5. Department of Biomedical Engineering, Cleveland Clinic, Cleveland, OH, 44106, USA.

11

12

13

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



15 **ABSTRACT**

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



35 INTRODUCTION

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

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



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



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

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



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

112

113

114 RESULTS AND DISCUSSION

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



125 democratize healthcare access, particularly in underserved regions, by providing affordable,
126 accurate, and accessible diagnostic solutions.^{14, 19, 54, 55}

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

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



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

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



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

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



193 95.52% sensitivity, 82.57% specificity, 90.72% accuracy, 90.3% precision, 92.8% F1 score,
194 79.95% MCC, and 97% AUC (**Supplementary Table 1**).

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

205 Our results indicated that DenseNet169, EfficientNetB5, and EfficientNetV2B0 exhibited
206 outstanding sensitivity values of 92.63%, 95.82%, and 91.93%, respectively (**Figure 3a and**
207 **Supplementary Figure 3-5**). ResNet50V2 (89.17%) and InceptionV3 (88.49%) demonstrated
208 high specificity values, while DenseNet169 displayed an exceptional specificity of 92.22%
209 (**Supplementary Table 2**). The confusion matrix revealed further insights into the performance
210 of these algorithms. DenseNet169 algorithm excelled in detecting negative samples, accurately
211 classifying 545 out of 591, while also achieving the second-highest performance in positive
212 classification with 930 out of 1004, resulting in the highest overall performance at 92% (**Figure**
213 **3b**). Other algorithms including EfficientNetB5 correctly identified 962 out of the tested 1004
214 positive samples. However, it misclassified 293 negative samples as positive, resulting in a 50.4%
215 performance rate for negative samples and an overall performance rate of 79%. EfficientNetV2B0



216 exhibited similar performance, albeit with a 7% overall performance rate downgrade, reflecting a
217 4% difference in true positive performance rate and an 11% decrease in true negative performance
218 rate. The results of MobileNetV2, EfficientNetV2B2, DenseNet201, InceptionV3, ResNet50V2,
219 and ResNet101V2 algorithms are shown in Supplementary **Figure 4, 5** with misclassification rates
220 $< 38\%$. The ROC analysis of the trained DL models, ResNet50V2 (AUC: 96%), ResNet101V2
221 (AUC: 96%), InceptionV3 (AUC: 95%) and DenseNet169 (AUC: 92%) and DenseNet201 (AUC:
222 90%) had the highest area under the ROC (**Supplementary Figure 6, 7**). Additionally, the
223 DenseNet169 model outperformed other models in terms of F1 score (93.94%) and accuracy
224 (92.48%) (**Supplementary Table 2**). Overall, DenseNet169 outperformed other models with the
225 performance metrics and gives the applicable model with 0.92 AUC (**Figure 3c**).

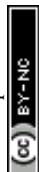
226 We compared the performance of Random Forest and DenseNet169, as these models had
227 outperformed others in our evaluations. To challenge them further, we used a set of 184 microchips
228 prepared with varying numbers of bubbles. A new test set of images was created under different
229 environmental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



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

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

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



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

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

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



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

297

298 **Conclusion**

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



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

317

318 **Data availability**

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

320

321 **Conflicts of interest**

322 There are no conflicts to declare.

323

324 **Acknowledgements**

325 The authors would like to thank Dr. Cyril R. A. John Chelliah, Ebenesh Chandrakumar,

326 Dheeksha Devaraj, and Aravinth Kumar for useful discussions and running initial trials.



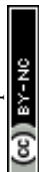
327 MATERIAL AND METHODS

328 Microfluidic chip model design and fabrication.

329 We developed a microfluid chip system that features a single microfluidic channel. The
330 microchip was designed using the vector graphics editor CorelDRAW Graphics suite software,
331 and fabricated of polymethyl methacrylate (PMMA) (3.125 mm thick), DSA film (100 μm thick,
332 3M, USA), and glass slides (25 mm x 75 mm). The fabrication process starts by cutting PMMA
333 and DSA film using a laser cutter (Boss Laser LS-1416, USA). The PMMA was prepared to
334 contain the microfluidic channel inlet and outlet, while DSA film included the main testing
335 channel. All materials were precleaned with 70% ethanol, and deionized water using lint-free
336 tissue. The surface of the cleaned glass slides was treated and cleaned using oxygen plasma (PE-
337 25, 100 mW, 15% oxygen; Plasma Etch Inc.) for 10 minutes. Then PMMA and DSA film were
338 assembled on the modified glass slide, forming the model microfluidic chip system. Each system
339 was loaded with platinum nanoparticle-seeded bubbles. PtNPs synthesized using our previously
340 published protocol were mixed with a peroxide-containing solution (5% hydrogen peroxide and
341 20% glycerol) and loaded on chip system. The concentration of added PtNPs was controlled to
342 prepare systems with variable numbers of bubbles (0 – >200 bubbles per chip), randomly
343 distributed within the microfluidic channel.

344 AI models selection, training and performance testing

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



349 models of MobileNetV2, EfficientNetV2B0, EfficientNetV2B2, DenseNet169, DenseNet201,
350 InceptionV3, ResNet50V2, EfficientNetB5 and ResNet101V2, were selected to support workflow
351 running on mobile devices and systems. We generated a dataset of 19,097 images of the model
352 microfluidic system captured using Moto XT1575, iPhone X and Vivo smartphones. The dataset
353 comprises two groups, i.e., positive (with > 10 bubbles per microchip) and negative (in range of <
354 10 bubbles per microchip) sample images. The microfluidic system imaging was performed at
355 different angles (0 – 360°) and backgrounds and environments to maximize the variations, and
356 make our dataset more robust and comprehensive. We used 15530 images for training, 1788
357 images for validation and 1012 images for testing the performance of the selected ML and DL
358 models in testing the model microfluidic system and classifying samples into positive and negative
359 based on bubble signal. We started the process by importing pre-trained models available from
360 Scikit-learn and Keras libraries to develop the selected ML and DL models, respectively. In the
361 pre-processing step, the images of our training dataset were resized to the input dimensions of the
362 selected models, leveraging the features learned by ImageNet pretrained network. We performed
363 the batch normalization then used Adam optimizer to fine-tune the network using a global learning
364 rate of 0.001. In addition, we employed a varied number of epochs to test the algorithms optimal
365 performance and we set the number to 50 epochs. Then we performed the transfer learning by
366 removing the final classification layer from the chosen networks and trained with our dataset. All
367 the algorithms were trained on Vector Workstation (Intel i9-10900x CPU and NVIDIA RTX
368 A6000 GPU, Lambda) and after training, we tested the performance of the best-performing ML
369 and DL algorithms individually using a challenging dataset of 400 images. This testing dataset
370 included rotated images, images with various colored backgrounds (matte, bright, reflective), and
371 images with lens distortion and brightness variations. The ML algorithms were evaluated using



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

376 **AI testing on a POC compatible system.**

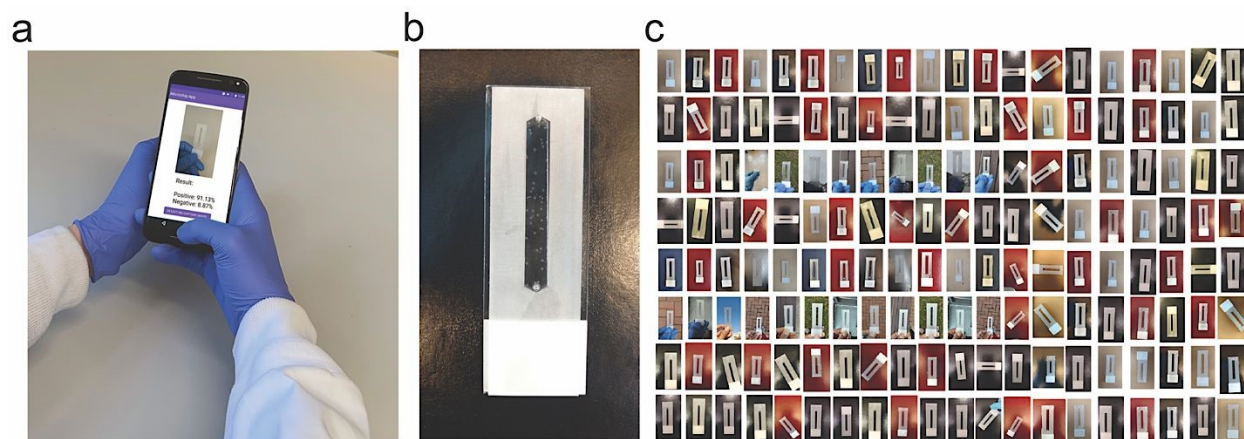
377 We utilized the open-source platform Android Studio (version Giraffe 2022.3.1) to develop
378 an AI-enabled mobile application. Android Studio offers an integrated development environment
379 (IDE) tailored for Android application development. The application facilitates the capture of
380 sensor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

392

393



394



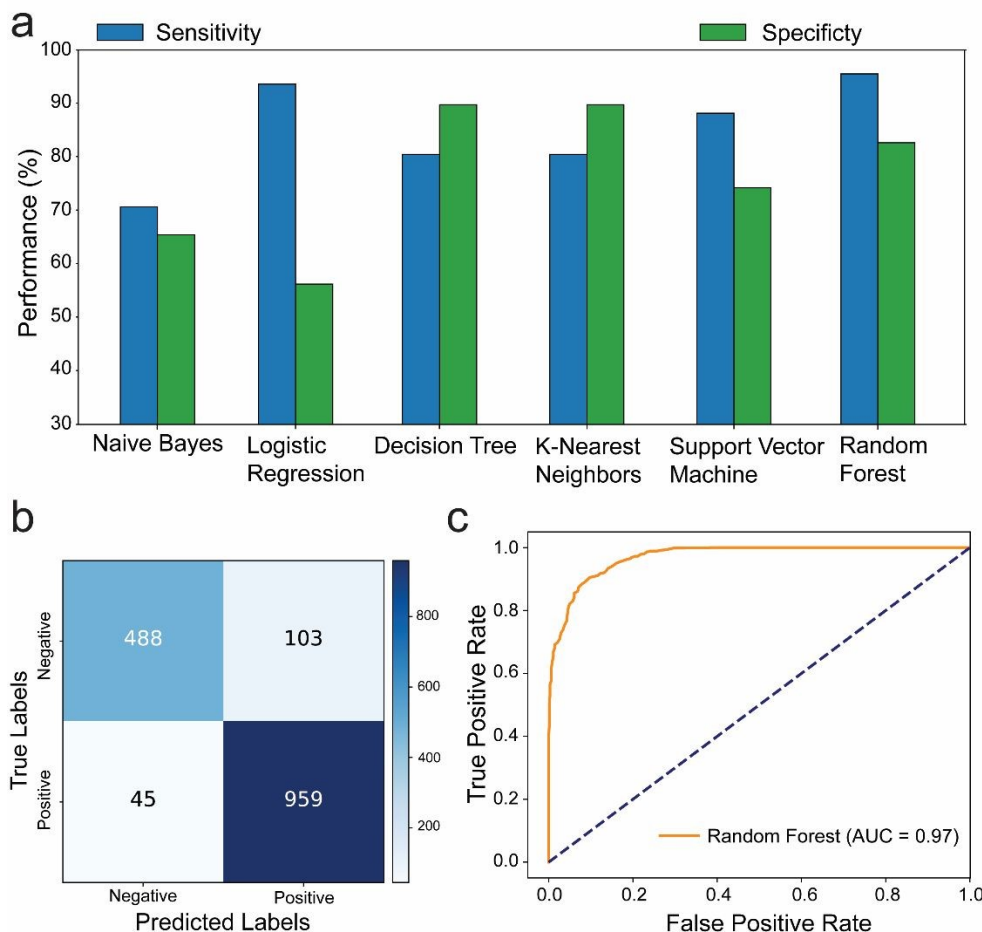
395

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

404



405



406

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

415

Open Access Article. Published on 20 September 2024. Downloaded on 10/2/2024 9:31:24 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.

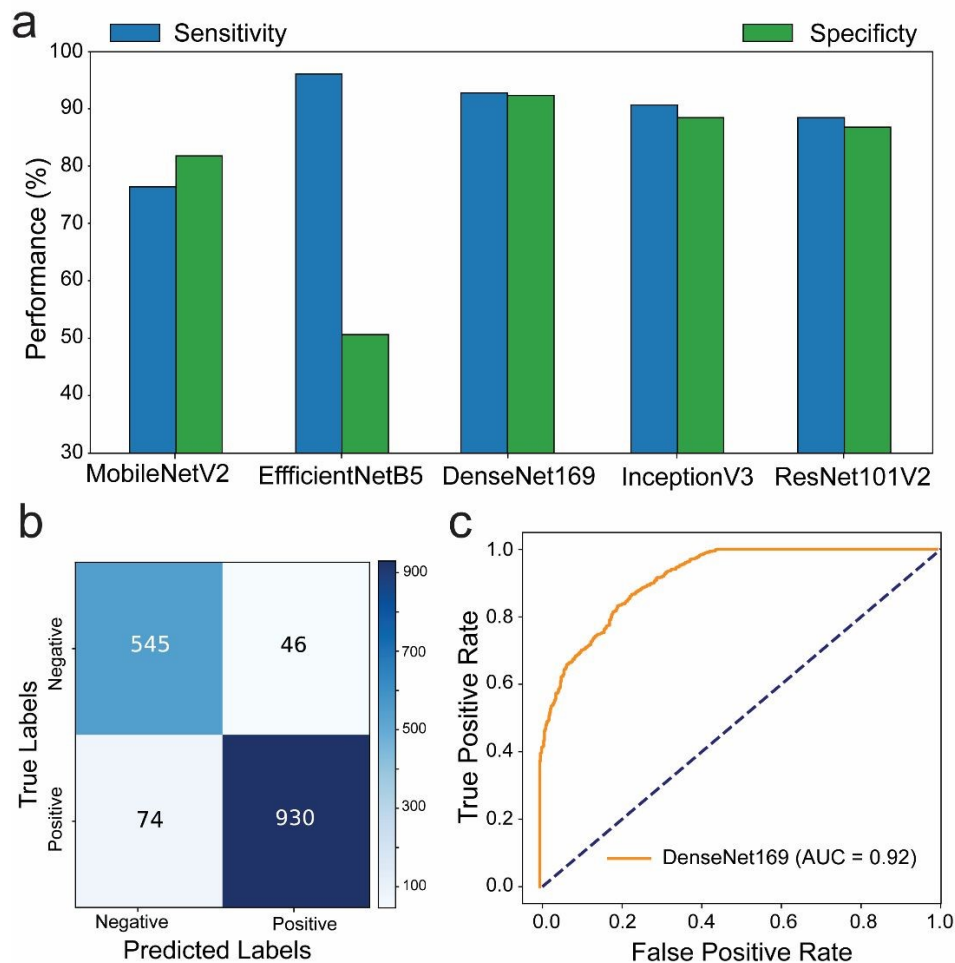


Lab on a Chip Accepted Manuscript

416

417

418

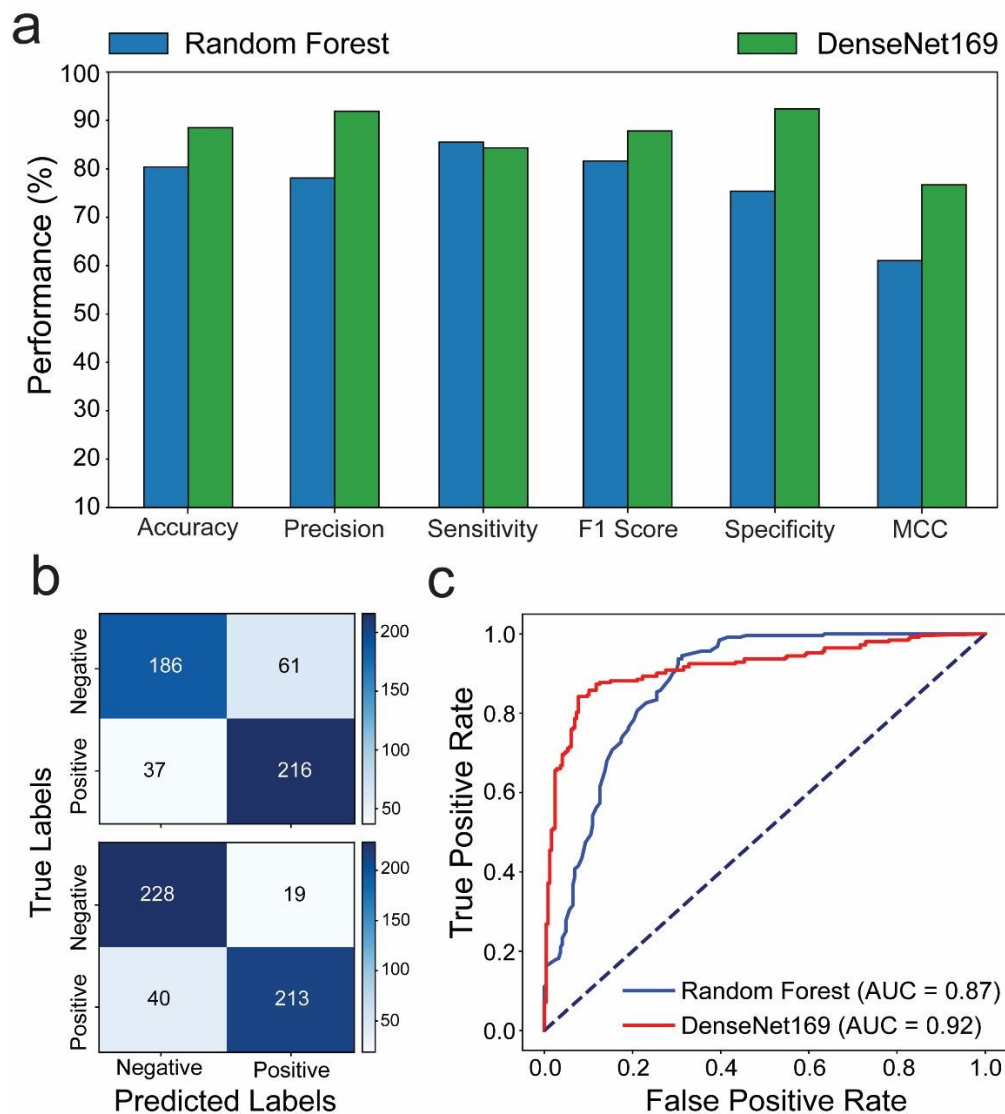


419

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

428





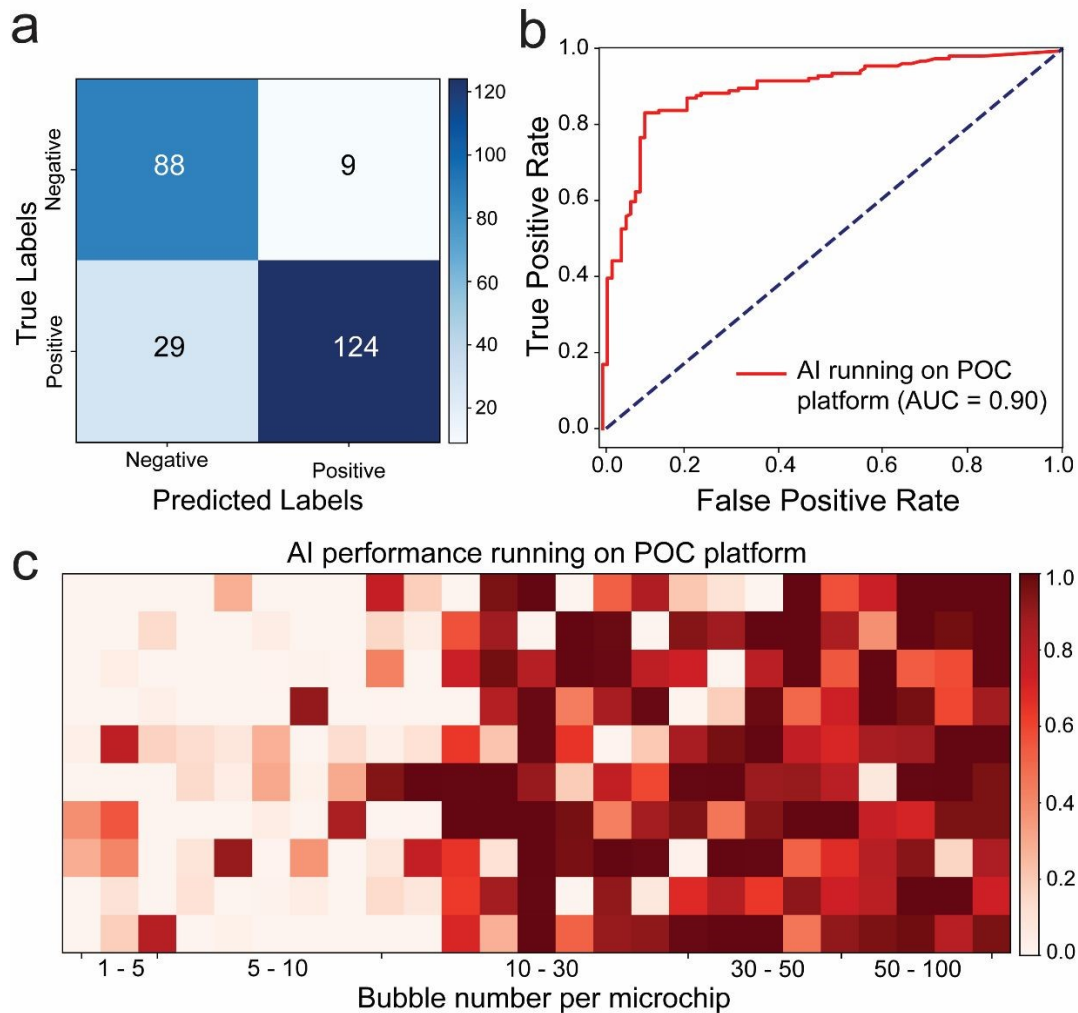
429

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

440



441



442

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

450

451

Lab on a Chip Accepted Manuscript



452 **References**

- 453
- 454 1. P. Rajpurkar, E. Chen, O. Banerjee and E. J. Topol, *Nature Medicine*, 2022, **28**, 31-38.
 - 455 2. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S.
456 Thrun and J. Dean, *Nature medicine*, 2019, **25**, 24-29.
 - 457 3. E. J. Topol, *Nature Medicine*, 2019, **25**, 44-56.
 - 458 4. J. N. Acosta, G. J. Falcone, P. Rajpurkar and E. J. Topol, *Nature Medicine*, 2022, **28**, 1773-1784.
 - 459 5. A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz and H. J. W. L. Aerts, *Nature Reviews*
460 *Cancer*, 2018, **18**, 500-510.
 - 461 6. D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang,
462 X. Wu and F. Yan, *cell*, 2018, **172**, 1122-1131. e1129.
 - 463 7. R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian
464 and A. Darzi, *npj Digital Medicine*, 2021, **4**, 65.
 - 465 8. P. Lambin, R. T. H. Leijenaar, T. M. Deist, J. Peerlings, E. E. C. de Jong, J. van Timmeren, S.
466 Sanduleanu, R. T. H. M. Larue, A. J. G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T.
467 Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger and S. Walsh,
468 *Nature Reviews Clinical Oncology*, 2017, **14**, 749-762.
 - 469 9. M. Sermesant, H. Delingette, H. Cochet, P. Jaïs and N. Ayache, *Nature Reviews Cardiology*, 2021,
470 **18**, 600-609.
 - 471 10. O. Oren, B. J. Gersh and D. L. Bhatt, *The Lancet Digital Health*, 2020, **2**, e486-e488.
 - 472 11. P. Yager, G. J. Domingo and J. Gerdes, *Annu. Rev. Biomed. Eng.*, 2008, **10**, 107-144.
 - 473 12. C. P. Y. Chan, W. C. Mak, K. Y. Cheung, K. K. Sin, C. M. Yu, T. H. Rainer and R. Renneberg, *Annual*
474 *review of analytical chemistry*, 2013, **6**, 191-211.
 - 475 13. C. Wang, M. Liu, Z. Wang, S. Li, Y. Deng and N. He, *Nano Today*, 2021, **37**, 101092.
 - 476 14. J. Riordon, D. Sovilj, S. Sanner, D. Sinton and E. W. Young, *Trends in biotechnology*, 2019, **37**,
477 310-324.
 - 478 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*
479 *Bioengineering*, 2023, **1**, 950-971.
 - 480 16. J. Zhou, J. Dong, H. Hou, L. Huang and J. Li, *Lab on a Chip*, 2024.
 - 481 17. W. Zhao, Y. Zhou, Y.-Z. Feng, X. Niu, Y. Zhao, J. Zhao, Y. Dong, M. Tan, Y. Xianyu and Y. Chen, *ACS*
482 *nano*, 2023, **17**, 13700-13714.
 - 483 18. Z. Ao, H. Cai, Z. Wu, L. Hu, A. Nunez, Z. Zhou, H. Liu, M. Bondesson, X. Lu and X. Lu, *Proceedings*
484 *of the National Academy of Sciences*, 2022, **119**, e2214569119.
 - 485 19. B. Wang, Y. Li, M. Zhou, Y. Han, M. Zhang, Z. Gao, Z. Liu, P. Chen, W. Du and X. Zhang, *Nature*
486 *Communications*, 2023, **14**, 1-18.
 - 487 20. H. Liu, L. Nan, F. Chen, Y. Zhao and Y. Zhao, *Lab on a Chip*, 2023, **23**, 2497-2513.
 - 488 21. J. Zheng, T. Cole, Y. Zhang, J. Kim and S.-Y. Tang, *Biosensors and Bioelectronics*, 2021, **194**,
489 113666.
 - 490 22. W. L. Bi, A. Hosny, M. B. Schabath, M. L. Giger, N. J. Birkbak, A. Mehrtash, T. Allison, O. Arnaout,
491 C. Abbosh and I. F. Dunn, *CA: a cancer journal for clinicians*, 2019, **69**, 127-157.
 - 492 23. N. Arora, A. K. Banerjee and M. L. Narasu, *Journal*, 2020, **15**, 717-724.
 - 493 24. K. P. Smith, H. Wang, T. J. Durant, B. A. Mathison, S. E. Sharp, J. E. Kirby, S. W. Long and D. D.
494 Rhoads, *Clinical Microbiology Newsletter*, 2020, **42**, 61-70.
 - 495 25. K. P. Smith and J. E. Kirby, *Clinical Microbiology and Infection*, 2020, **26**, 1318-1323.
 - 496 26. A. J. London, *Hastings Center Report*, 2019, **49**, 15-21.
 - 497 27. C. Gilvary, N. Madhukar, J. Elkhader and O. Elemento, *Trends in pharmacological sciences*, 2019,
498 **40**, 555-564.



- 499 28. J. E. Dayhoff and J. M. DeLeo, *Cancer: Interdisciplinary International Journal of the American*
500 *Cancer Society*, 2001, **91**, 1615-1635.
- 501 29. O. Koteluk, A. Wartecki, S. Mazurek, I. Kołodziejczak and A. Mackiewicz, *Journal of Personalized*
502 *Medicine*, 2021, **11**, 32.
- 503 30. H. Bhaskar, D. C. Hoyle and S. Singh, *Computers in biology and medicine*, 2006, **36**, 1104-1125.
- 504 31. A. A. de Hond, A. M. Leeuwenberg, L. Hooft, I. M. Kant, S. W. Nijman, H. J. van Os, J. J. Aardoom,
505 T. P. Debray, E. Schuit and M. van Smeden, *NPJ digital medicine*, 2022, **5**, 2.
- 506 32. M. Sumner, E. Frank and M. Hall, 2005.
- 507 33. H. M. Sani, C. Lei and D. Neagu, 2018.
- 508 34. X. Zheng, J. Jia, S. Guo, J. Chen, L. Sun, Y. Xiong and W. Xu, *IEEE Journal of Selected Topics in*
509 *Applied Earth Observations and Remote Sensing*, 2021, **14**, 2222-2235.
- 510 35. Z. Zheng, 1998.
- 511 36. L. Bottou and C.-J. Lin, 2007.
- 512 37. K. Hajebi, Y. Abbasi-Yadkori, H. Shahbazi and H. Zhang, 2011.
- 513 38. Y. Zhang, L. Wang, J. Zhao, X. Han, H. Wu, M. Li and M. Deveci, *Information Sciences*, 2024, **670**,
514 120644.
- 515 39. R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A.
516 Bosselut and E. Brunskill, *arXiv preprint arXiv:2108.07258*, 2021.
- 517 40. R. C. Deo, *Circulation*, 2015, **132**, 1920-1930.
- 518 41. A. Alanazi, *Informatics in Medicine Unlocked*, 2022, **30**, 100924.
- 519 42. M. A. Morid, A. Borjali and G. Del Fiol, *Computers in biology and medicine*, 2021, **128**, 104115.
- 520 43. A. W. Salehi, S. Khan, G. Gupta, B. I. Alabdullah, A. Almjally, H. Alsolai, T. Siddiqui and A. Mellit,
521 *Sustainability*, 2023, **15**, 5930.
- 522 44. M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu and X.
523 Zhao, *Journal of X-ray Science and Technology*, 2020, **28**, 821-839.
- 524 45. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, *Proceedings of the IEEE*, 2020,
525 **109**, 43-76.
- 526 46. S. J. Pan and Q. Yang, *IEEE Transactions on knowledge and data engineering*, 2009, **22**, 1345-
527 1359.
- 528 47. J. Ooge, G. Stiglic and K. Verbert, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge*
529 *Discovery*, 2022, **12**, e1427.
- 530 48. K. Moulaei, A. Yadegari, M. Baharestani, S. Farzanbakhsh, B. Sabet and M. R. Afrash,
531 *International Journal of Medical Informatics*, 2024, 105474.
- 532 49. X. Tang, *BJR| Open*, 2019, **2**, 20190031.
- 533 50. N. Hasani, M. A. Morris, A. Rahmim, R. M. Summers, E. Jones, E. Siegel and B. Saboury, *PET*
534 *clinics*, 2022, **17**, 1-12.
- 535 51. Z. Angehrn, L. Haldna, A. S. Zandvliet, E. Gil Berglund, J. Zeeuw, B. Amzal, S. A. Cheung, T. M.
536 Polasek, M. Pfister and T. Kerbusch, *Frontiers in pharmacology*, 2020, **11**, 759.
- 537 52. Y. Yang, F. Xu, J. Chen, C. Tao, Y. Li, Q. Chen, S. Tang, H. K. Lee and W. Shen, *Biosensors and*
538 *Bioelectronics*, 2023, 115233.
- 539 53. D. Xu, X. Huang, J. Guo and X. Ma, *Biosensors and Bioelectronics*, 2018, **110**, 78-88.
- 540 54. D. McIntyre, A. Lashkaripour, P. Fordyce and D. Densmore, *Lab on a Chip*, 2022, **22**, 2925-2937.
- 541 55. I. Hernández-Neuta, F. Neumann, J. Brightmeyer, T. Ba Tis, N. Madaboosi, Q. Wei, A. Ozcan and
542 M. Nilsson, *Journal of internal medicine*, 2019, **285**, 19-39.
- 543 56. Y. Jiao and P. Du, *Quantitative Biology*, 2016, **4**, 320-330.
- 544 57. I. Stančin and A. Jović, 2019.
- 545 58. B. J. Erickson, P. Korfiatis, Z. Akkus, T. Kline and K. Philbrick, *Journal of digital imaging*, 2017, **30**,
546 400-405.



547

Open Access Article. Published on 20 September 2024. Downloaded on 10/2/2024 9:31:24 PM.
This article is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported Licence.



Lab on a Chip Accepted Manuscript

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

Mert Tunca Doganay¹, Purbali Chakraborty¹, Sri Moukthika¹, Soujanya Jammalamadaka¹,

Dheerendranath Battalapalli¹, 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



Data Availability

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

