

Assessing Automated Machine Learning Services for COVID-19 Detection from X-Ray and CT Images

A Real-Time Smartphone Application Case Study

Evaluating Low-Resource AI Diagnostic Tools for Equitable Pandemic Response in Rural and Underserved Healthcare Settings

Razib Mustafiz^{1*}, Khaled Mohsin, MD²

¹ School of Computing, Dublin City University, Ireland | ² Department of Cardiology, St James's Hospital, Trinity College Dublin, Ireland

* Corresponding author: mohammad.mustafiz2@mail.dcu.ie

Article Info	ABSTRACT
Manuscript Status: Retrospective, methodological case study built on publicly available chest-imaging datasets. Prospective, demographically diverse clinical validation is identified by the authors as necessary future work. Keywords: COVID-19; machine learning; AutoML; smartphone diagnostics; deep learning; generative adversarial networks; global health equity	Background: The 2020 SARS-CoV-2 pandemic created urgent demand for non-interventional, scalable AI tools to support pulmonary disease screening, at a time when radiologists were newly confronting an unfamiliar disease presentation. Objective: This study evaluated Microsoft Cognitive Service (Custom Vision), an automated machine learning (AutoML) platform, for detecting and classifying COVID-19-induced pneumonia against other viral/bacterial pneumonia and normal lung tissue from chest X-ray and CT images — with a specific focus on whether a cloud-trained model could be ported to run real-time, offline inference on a smartphone. Methodology: Using publicly available, curated X-ray and CT datasets (COVID-19-positive, normal, pneumonic, and a control “inapplicable” household-objects class), the authors ran seven training iterations in Custom Vision’s Rapid Application Development environment, varying class counts and balance, then deployed the best-performing model both as a REST API-backed web application and as an on-device TensorFlow Lite / Core ML mobile application for Android and iOS. Key Findings: The optimized model (Iteration 6/7) achieved 96.8% average precision as a web application, with 100% recall and 99.4% average precision specifically for the COVID-19 class. However, the identical trained network performed markedly worse once ported to a smartphone for real-time, on-device inference — average precision for some classes (e.g., Pneumonia) fell from the high-90s on the web to as low as 15% on mobile — a gap the authors identify as the study’s central and most policy-relevant finding. Policy Impact: The web/mobile performance gap, and the underlying COVID-19 data scarcity and class imbalance that likely drive it, show that cloud-validated diagnostic AI cannot be assumed to perform equivalently on-device — which matters directly for rural and resource-constrained health systems that depend on offline, smartphone-based screening where internet connectivity is unreliable.

Institutional publication of the Centre for Humanitarian Research and Social Development (CHRSD Foundation), Research Paper Series.

1. INTRODUCTION AND CONTEXT

1.1 Problem Statement and Background

Lung disease has long been a major contributor to global morbidity and mortality, with lung cancer historically the predominant concern. The emergence of SARS-CoV-2 added an acute, fast-moving respiratory threat on top of this baseline burden, and because the virus’s radiographic presentation was unfamiliar, radiologists were frequently not yet equipped to reliably distinguish COVID-19-induced pneumonia from ordinary viral or bacterial pneumonia on chest X-ray alone, without the patient’s symptoms and history in hand.

Since 2016, cloud providers have progressively democratized machine learning through Automated ML (AutoML) and Rapid Application Development (RAD) platforms — Microsoft Custom Vision, Google Cloud AutoML, and Apple Create ML — that let non-specialists build transfer-learning-based image classifiers by uploading and tagging images in a graphical environment, without deep learning expertise. This study uses Microsoft Custom Vision, built on a pre-trained convolutional neural network, to test whether this democratized tooling could be turned into a credible, low-resource COVID-19 screening aid, and specifically whether it could run entirely on a smartphone, with zero network dependency.

1.2 Objectives and Scope

Primary Objective: To build and evaluate a smartphone-based, on-device real-time inference system that detects COVID-19-induced pneumonia from chest X-ray and CT images without requiring an internet connection — intended for rural and under-resourced settings where connectivity is poor or absent.

Secondary Objective: To build and evaluate a parallel web-portal implementation of the same trained model, served over a REST API, as a comparison point for the mobile deployment.

The analysis is explicitly a retrospective, methodological case study built on publicly available datasets rather than a clinical validation trial; the authors frame prospective, demographically diverse clinical validation as necessary future work rather than something this paper establishes.

1.3 Relevance to Humanitarian and Social Development Goals

The central motivation is equity of access: AI-assisted primary screening that runs entirely on a smartphone’s own processor, with no internet dependency, can bring a basic diagnostic capability to rural and understaffed health centers in developing countries that have an X-ray machine but unreliable connectivity. With global smartphone penetration already above 45% in 2020, the authors argue a validated on-device model could reach a very large underserved population directly. This aligns with:

SDG	Relevance
SDG 3 – Good Health and Well-Being	Expands access to primary pandemic screening in under-resourced health systems
SDG 9 – Industry, Innovation and Infrastructure	Tests whether commodity smartphone hardware can substitute for specialist diagnostic infrastructure
SDG 10 – Reduced Inequalities	Targets the connectivity and resource gap between well-equipped and rural/under-resourced clinics

The study’s central finding — that the same model performed far worse on-device than in the cloud — is itself a humanitarian caution: it shows that equity-motivated tools cannot be assumed to work as intended simply because they run offline, and that on-device validation is not optional.

2. METHODOLOGY AND ANALYTICAL FRAMEWORK

2.1 Data Collection and Preparation

Images were drawn from publicly available, curated sources: COVID-19-positive chest X-rays, normal and other-pneumonia chest X-rays, COVID-19 and normal CT scans, and a household-objects set used to train an “Inapplicable” control class. Most chest radiographs were standard postero-anterior (PA) views.

Dataset	Train	Test	Validation
COVID-19 X-Ray	295	—	25
Normal X-Ray	1,341	234	8
Pneumonia X-Ray	3,875	390	8
Normal CT	397	—	—
COVID-19 CT	349	—	—
Household objects (“Inapplicable”)	185	123	—

Custom Vision recommends at least 50 images per tag and roughly equal counts across classes; the COVID-19 X-ray class (295 images) was the smallest available, meaning any model using the full Pneumonia and Normal sets would be structurally imbalanced. Custom Vision does not require a separately supplied test set — it allocates held-out test data automatically from the uploaded images — so the authors used the labeled test partitions here for independent, manual performance review.

2.2 Data Loading and Model Building

The pipeline used Microsoft Custom Vision’s paired Training and Prediction resources in Azure. Seven training iterations were run with different combinations of classes and image counts, using Custom Vision’s four pre-trained model families across two domains — **Compact** (exportable for on-device mobile/edge inference) and **General** (cloud/REST-API inference only) — and two training modes, **Quick** and **Advanced** (the latter tuned for harder, fine-grained, less-augmented data). Iterations 1–6 used Advanced training on the Compact domain; Iteration 7 repeated Iteration 6’s data on the General domain for REST API publication.

Across the seven iterations, the authors progressively added classes and then corrected for class imbalance: Iteration 1 used only two balanced classes (25 images each, COVID-19-positive vs. negative) and reached 100% on every metric, but was judged impractical given its small size and binary scope. Iterations 2–5 added more classes (Inapplicable, Pneumonia, Normal, and CT variants) with uneven image counts; Custom Vision itself flagged the resulting imbalance, and these models proved unstable when ported for mobile inference. Iteration 6 fixed this by using exactly 275 images per class across six classes (Inapplicable, COVID-19, Pneumonia, Normal, COVID-19 CT, Normal CT) and produced the first stable mobile deployment; Iteration 7 reran the same balanced dataset on the General domain to obtain REST API access for the web application.

2.3 Implementation

Web Application: Built in PHP, calling the Custom Vision REST API published from Iteration 7. The authors note a data-privacy consideration: images uploaded through the live REST API are stored on the Custom Vision cloud service and can be reused for further training, which raises a jurisdiction-dependent data-protection question. They note that exporting the trained network in TensorFlow.js format and running inference locally in the browser avoids this, since no image data would leave the client.

Mobile Application: Built for Android and iOS using the Iteration 6 model exported in TensorFlow Lite and Core ML format respectively, enabling fully offline, on-device inference with no network dependency — the paper’s central use case. Both apps are publicly available (Android/iOS source repositories and an APK download are listed in the original source; *editorial flag: verify these third-party links are still live before this paper is published, as they were not independently checked for this adaptation*).

2.4 Evaluation Framework

Model performance was assessed with the standard accuracy / precision / recall / F1 suite, using true positive (A), false negative (B), false positive (C), and true negative (D) counts ($\text{Accuracy} = (A+D)/(A+B+C+D)$). The authors additionally varied the inference-stage **probability threshold T** (the cutoff controlling the precision/recall trade-off) across 50%, 40%, 30%, 20%, and 10%, to characterize how sensitive the deployed model’s error profile is to this single tunable parameter — directly relevant to a screening context where missed positives (false negatives) are more costly than false alarms.

3. KEY FINDINGS AND ANALYSIS

3.1 Best Cloud-Side Model: Iteration 6/7 Performance

The balanced, six-class Iteration 6/7 model (275 images per class) was the strongest and most stable configuration, achieving 96.1% precision, 96.1% recall, and 96.8% average precision overall when served as a web application — with 100% recall and 99.4% average precision specifically for the COVID-19 class. This is the headline cloud-side result, and on its own would support cautious optimism about AutoML for this task.

3.2 The Central Finding: Web-to-Mobile Performance Collapse

The same trained network performed markedly worse once ported to a smartphone for on-device inference — the paper’s central and most policy-relevant result:

Class	Web Application (Avg. Precision %)	Mobile Application (Avg. Precision %)
Inapplicable	100	100
Covid-19	96.5	87
Normal	94.7	25
Normal CT	94.1	20
Covid-19 CT	89.7	40
Pneumonia	98.1	15

Only the “Inapplicable” control class held steady; every clinical class degraded substantially on-device, with Pneumonia falling from 98.1% to 15%. The authors attribute this to a combination of insufficient COVID-19 training data, residual class imbalance, and smartphone hardware variance (newer, more powerful Android devices performed better in their informal testing of six handsets), but state plainly that the magnitude and root cause of the web-to-mobile gap was not fully explained and needs further study.

Editorial note on the data: The source document also reports confidence intervals around these mobile/web comparisons (its Table 5) with several class-level lower bounds below 0% and upper bounds above 100% — not meaningful for a percentage metric. This is very likely a computational or transcription artifact in the source rather than a real finding, and

CHRSD recommends the original confidence-interval calculation be re-run and corrected before this comparison is cited in a policy context. The underlying mean performance figures in the table above are reproduced as reported.

3.3 Probability Threshold (T) Trade-off

Varying the inference threshold T on the deployed model produced a consistent precision/recall trade-off while average precision stayed flat at 96.8%:

Threshold T	Precision %	Recall %	Average Precision %
50%	96.1	96.1	96.8
40%	95.8	96.1	96.8
30%	95.5	96.4	96.8
20%	93.0	97.3	96.8
10%	91.2	97.6	96.8

As T falls from 50% to 10%, precision drops from 96.1% to 91.2% while recall rises from 96.1% to 97.6% — i.e., the false-negative rate falls to roughly 2.4%. Because a missed COVID-19 case (false negative) carries a much higher cost in a screening context than a false alarm (which a subsequent RT-PCR test can resolve), the authors recommend operating at a low threshold (around 10%) to minimize missed cases, accepting more false positives in exchange.

3.4 Class Imbalance as the Recurring Driver

Across iterations 2–5, Custom Vision itself detected and flagged class-count imbalance, and these models were correspondingly unstable on mobile. Only once image counts were equalized at 275 per class (Iteration 6) did the mobile deployment stabilize. This points to COVID-19 X-ray data scarcity — the smallest class throughout — as the structural constraint shaping every other result in this study, including the web/mobile gap in 3.2.

3.5 Implementation-Level Observations

Uploading images through the live Custom Vision REST API stores them on Microsoft’s cloud service, a data-privacy consideration for any future clinical deployment; the authors note that a TensorFlow.js in-browser export avoids this by keeping inference (and the data) client-side. Separately, informal testing across six Android handsets suggested that devices issued after 2018, with faster processors and more RAM, handled on-device inference more reliably — the iPhone build was not separately tested in this study.

4. STRATEGIC IMPLICATIONS AND POLICY RECOMMENDATIONS

4.1 Require On-Device Validation, Not Just Cloud Validation

The study’s central lesson is that a model’s cloud/web performance does not predict its on-device performance. Health ministries, digital-health programmes, and regulators evaluating any smartphone-based AI diagnostic tool should require separate, documented validation of the actual deployed mobile artifact — not only of the cloud-trained model it was exported from — before any field or pilot deployment.

Stakeholders: Ministries of Health, national digital-health/e-health agencies, medical device regulators. Timeline: incorporate as a standing requirement in any AI-diagnostics procurement or pilot-approval checklist within 12 months.

4.2 Treat Reference Data Scarcity as the Priority Investment

Class imbalance — driven by the small COVID-19 X-ray dataset relative to Normal and Pneumonia data — was the recurring cause of both training instability and (plausibly) the mobile performance collapse. Funders and research institutions should prioritize building and openly depositing larger, balanced, diverse chest-imaging datasets for emerging pathogens, rather than assuming AutoML transfer learning can fully compensate for data scarcity.

Stakeholders: Research funders, national radiology/imaging societies, open data repositories (e.g., GEO-equivalent imaging archives). Timeline: dataset-investment call within 12 months of any future respiratory outbreak.

4.3 Set Screening Thresholds to Minimize False Negatives, Explicitly

Where such tools are used for primary pandemic screening ahead of confirmatory RT-PCR testing, the probability threshold should be deliberately set low (the study suggests around 10%) to minimize missed cases, with the resulting higher false-positive rate treated as an acceptable, lower-cost trade-off resolved by the follow-on test — and this threshold choice should be stated explicitly in any deployment, not left at a tool's untuned default.

Stakeholders: Clinical AI deployment teams, health ministries, device manufacturers. Timeline: immediate, at point of deployment.

4.4 Address Data Governance Before Field Deployment

The authors flag that images uploaded through a live Custom Vision REST API are retained on the cloud vendor's servers — a data-protection question that will fall differently under different jurisdictions' health-data law. Any deployment handling real patient images should resolve data residency and consent questions beforehand, and should prefer architectures (such as in-browser or on-device inference) that keep patient images off third-party cloud infrastructure entirely.

Stakeholders: Data protection authorities, health ministries, implementing NGOs and hospitals. Timeline: resolved before any pilot involving real patient data.

4.5 Fund Synthetic Data Augmentation Research, With Ethical Guardrails

The authors propose Generative Adversarial Networks (DCGAN, PGGAN) as a route to augment scarce COVID-19 imaging data, while explicitly flagging the ethical and legal questions raised by synthetic medical images (e.g., provenance, consent, and downstream clinical use of AI-generated rather than real patient images). Funders supporting this direction should require upfront ethical review rather than treating it as a purely technical augmentation step.

Stakeholders: Research funders, bioethics review boards, AI/ML research groups. Timeline: ethical framework developed alongside, not after, any GAN-augmentation research programme.

5. CONCLUSION

This study demonstrates that automated, commercially available machine learning platforms can be used to rapidly build a credible COVID-19 screening classifier from publicly available chest imaging data, reaching 96.8% average precision as a web application. Its more important contribution, however, is a cautionary one: the identical trained network did not perform equivalently once ported for real-time, offline inference on a smartphone, with several classes losing the majority of their average precision on-device. For a field whose

central promise is exactly this kind of offline, low-resource deployment, that gap is the finding that matters most.

Key takeaways:

- Cloud/web validation metrics cannot stand in for on-device validation — the two can diverge sharply even with an identical underlying model.
- Reference data scarcity and class imbalance (here, limited COVID-19 imaging data) is a recurring, upstream driver of downstream instability.
- Threshold tuning is a low-cost, high-leverage lever for screening tools where false negatives are costlier than false positives.
- Data governance (where patient images are stored and processed) needs resolving before, not after, clinical deployment.
- Smartphone-based diagnostic AI remains a genuinely promising route to equitable health access, provided the web-to-mobile performance gap identified here is investigated and closed.

Vision for future directions: With global smartphone penetration already above 45% in 2020, a properly validated on-device diagnostic tool could reach a very large underserved population. The authors identify generative data augmentation (DCGAN, PGGAN) to address COVID-19 data scarcity, systematic study of smartphone hardware variance, and formal clinical validation trials as the next steps; CHRSD would add, from this adaptation's own review, a dedicated root-cause investigation of the web-to-mobile performance gap itself (Section 3.2) as a prerequisite before any such tool is recommended for field use.

Limitations: This is a retrospective, methodological case study built entirely on publicly available datasets, not a prospective clinical trial on demographically diverse patients; the authors note the lack of published data on ethnically comparable populations prevented broader comparison. The root cause of the web-to-mobile performance gap was not established. The iPhone build was not independently tested. The reported confidence intervals around the mobile/web comparison (Table 5 in Section 3.2) contain implausible bounds outside the 0–100% range and should be recalculated before being cited.

REFERENCES (APA 7TH EDITION)

- Abbott, L. M., & Smith, S. D. (2018). Smartphone apps for skin cancer diagnosis: Implications for patients and practitioners. *Australasian Journal of Dermatology*, 59(3), 168–170.
- Alphabet Inc. (n.d.). *Cloud AutoML*. Retrieved May 2020, from <https://cloud.google.com/automl/>
- Apple Inc. (n.d.). *Create ML*. Retrieved May 2020, from <https://developer.apple.com/documentation/createml>
- Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D.-Y., Chen, L., & Wang, M. (2020). Presumed asymptomatic carrier transmission of COVID-19. *JAMA*.
- Bhat, R., Hamid, A., Kunin, J. R., et al. (2020). Chest imaging in patients hospitalized with COVID-19 infection — A case series. *Current Problems in Diagnostic Radiology*. <https://doi.org/10.1067/j.cpradiol.2020.04.001>
- Bloice, M. D., Roth, P. M., & Holzinger, A. (2019). Biomedical image augmentation using Augmentor. *Bioinformatics*, 35(21), 4522–4524. <https://doi.org/10.1093/bioinformatics/btz259>
- Borkowski, A., Viswanadham, N. A., Thomas, L. B., et al. (2020). Using artificial intelligence for COVID-19 chest X-ray diagnosis. <https://doi.org/10.1101/2020.05.21.20106518>

- Borkowski, A. A., Wilson, C. P., Borkowski, S. A., et al. (2019). Comparing artificial intelligence platforms for histopathologic cancer diagnosis. *Federal Practitioner*, 36(10), 456–463.
- Caltech Vision Group. (n.d.). *Home objects dataset*. Retrieved May 2020, from http://www.vision.caltech.edu/pmoreels/Datasets/Home_Objects_06/
- Cepheid. (2020). *Xpert® Xpress SARS-CoV-2 has received FDA Emergency Use Authorization*. Retrieved May 10, 2020, from <https://www.cepheid.com/coronavirus>
- Cherian, T., Mulholland, E. K., Carlin, J. B., et al. (2005). Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bulletin of the World Health Organization*, 83, 353–359.
- Cohen, J. P., Morrison, P., & Dao, L. (2020). COVID-19 image data collection. *arXiv:2003.11597*. <https://github.com/ieee8023/covid-chestxray-dataset>
- Contreras-Naranjo, J. C., Wei, Q. S., & Ozcan, A. (2016). Mobile phone-based microscopy, sensing, and diagnostics. *IEEE Journal of Selected Topics in Quantum Electronics*, 22.
- Dallet, C., Kareem, S., & Kale, I. (2014). Real time blood image processing application for malaria diagnosis using mobile phones. *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2405–2408. <https://doi.org/10.1109/ISCAS.2014.6865657>
- De Fauw, J., Ledsam, J. R., Romera-Paredes, B., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9), 1342–1350.
- Diederich, B., Wartmann, R., Schadwinkel, H., et al. (2018). Using machine-learning to optimize phase contrast in a low-cost cellphone microscope. *PLoS ONE*, 13(3), e0192937.
- Fang, Y., Zhang, H., Xie, J., et al. (2020). Sensitivity of chest CT for COVID-19: Comparison to RT-PCR. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7)
- Franquet, T. (2001). Imaging of pneumonia: Trends and algorithms. *European Respiratory Journal*, 18(1), 196–208.
- Gallegos, D., Long, K. D., Yu, H., et al. (2013). Label-free biodetection using a smartphone. *Lab on a Chip*, 13, 2124–2132.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani et al. (Eds.), *Advances in Neural Information Processing Systems (NIPS 2014)* (Vol. 2, pp. 2672–2680). MIT Press.
- Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., Bernheim, A., & Siegel, E. (2020). Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. *arXiv:2003.05037*.
- Han, C., Hayashi, H., Rundo, L., et al. (2018). GAN-based synthetic brain MR image generation. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 734–738. <https://doi.org/10.1109/ISBI.2018.8363678>
- Han, S. S., Park, G. H., Lim, W., et al. (n.d.). Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. [editorial flag: year and venue not stated in source document — verify before publication.]
- He, K., Zhang, X., Ren, S., & Sun, J. (n.d.). Deep residual learning for image recognition. [editorial flag: the source gives no year or venue; this is very likely He et al. (2016), IEEE CVPR, but verify before citing.]
- Hernández-Neuta, I., Neumann, F., Brightmeyer, J., et al. (n.d.). Smartphone-based clinical diagnostics: Towards democratization of evidence-based health care. [editorial flag: year and venue not stated in source document — verify before publication.]
- Italian Society of Medical and Interventional Radiology. (2020). [Case report, March 31, 2020]. Retrieved May 2020, from <https://www.sirm.org/en/2020/03/31/covid-19-case-4/>
- Jacobi, A., Chung, M., Bernheim, A., & Eber, C. (2020). Cardiothoracic imaging: Portable chest X-ray in coronavirus disease-19 (COVID-19) — A pictorial review. <https://doi.org/10.1016/j.clinimag.2020.04.001>

- Jayatilake, D., Suzuki, K., Teramoto, Y., et al. (2014). Swallowscope: A smartphone based device for the assessment of swallowing ability. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, 697–700. <https://doi.org/10.1109/BHI.2014.6864459>
- Jiang, F., Jiang, Y., Zhi, H., et al. (2017). Artificial intelligence in health care: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243.
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of GANs for improved quality, stability, and variation. *arXiv:1710.10196*.
- Kazemini, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., & Mukhopadhyay, A. (2018). GANs for medical image analysis. *arXiv:1809.06222*.
- Kermany, D., Goldbaum, M., Cai, W., Lewis, M. A., Xia, H., & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172, 1122–1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
- Kovalev, V., & Kazlouski, S. (n.d.). Examining the capability of GANs to replace real biomedical images in classification models training. [editorial flag: year and venue not stated in source document — verify before publication.]
- Kovalev, V., Liauchuk, V., Kalinovskiy, A., & Shukelovich, A. (2017). Comparison of conventional and deep learning methods of image classification on a database of chest radiographs. *International Journal of Computer Assisted Radiology and Surgery*, 12, 139–140.
- Langs, G., Röhrich, S., Hofmanninger, J., et al. (2018). Machine learning: From radiomics to discovery and routine. *Der Radiologe*. <https://doi.org/10.1007/s00117-018-0407-3>
- Li, Y., & Shen, L. (n.d.). Skin lesion analysis towards melanoma detection using deep learning network. [editorial flag: year and venue not stated in source document — verify before publication.]
- Lipowicz, M., & Garcia, A. (2015). Handheld device adapted to smartphone cameras for the measurement of sodium ion concentrations at saliva-relevant levels via fluorescence. *Bioengineering*, 2, 122–138.
- Lomoro, P., Verde, F., Zerboni, F., et al. (2020). COVID-19 pneumonia manifestations at the admission on chest ultrasound, radiographs, and CT: Single-center study and comprehensive radiologic literature review. <https://doi.org/10.1016/j.ejro.2020.100231>
- Mahbod, A., Ecker, R., & Ellinger, I. (2017). Skin lesion classification using hybrid deep neural networks. *arXiv:1702.08434*.
- Martinez, A. W., Phillips, S. T., Carrilho, E., Thomas, S. W., III, Sindi, H., & Whitesides, G. M. (2008). Simple telemedicine for developing regions: Camera phones and paper-based microfluidic devices for real-time, off-site diagnosis. *Analytical Chemistry*, 80, 3699–3707.
- Mauk, M. G. (n.d.). Calling in the test: Smartphone-based urinary sepsis diagnostics. <https://doi.org/10.1016/j.ebiom.2018.09.001>
- McBee, M. P., Awan, O. A., Colucci, A. T., et al. (2018). Deep learning in radiology. *Academic Radiology*. <https://doi.org/10.1016/j.acra.2018.02.018>
- Microsoft Corporation. (n.d.). *Microsoft Custom Vision*. Retrieved May 2020, from <https://azure.microsoft.com/en-us/services/cognitive-services/custom-vision-service/>
- Microsoft Corporation. (n.d.). *Use your model with the prediction API*. Retrieved May 2020, from <https://docs.microsoft.com/en-us/azure/cognitive-services/custom-vision-service/use-prediction-api>
- Monkam, P., Qi, S., Xu, M., et al. (2018). CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images. *BioMedical Engineering OnLine*, 17(1), 96.
- Mooney, P. T. (n.d.). *Chest X-ray images (pneumonia)*. Kaggle. Retrieved May 2020, from <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>
- Mudanyali, O., Dimitrov, S., Sikora, U., Padmanabhan, S., Navruz, I., & Ozcan, A. (2012). Integrated rapid-diagnostic-test reader platform on a cellphone. *Lab on a Chip*, 12, 2678–2686.

- Orth, A., Wilson, E. R., Thompson, J. G., & Gibson, B. C. (2018). A dual-mode mobile phone microscope using the onboard camera flash and ambient light. *Scientific Reports*, 8, 3298.
- Ozcan, A. (2014). Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab on a Chip*, 14, 3187–3194.
- Panahiazar, M., Taslimitehrani, V., Pereira, N., et al. (n.d.). Using EHRs and machine learning for heart failure survival analysis. [editorial flag: year and venue not stated in source document — verify before publication.]
- Poostchi, M., Silamut, K., Maude, R. J., et al. (n.d.). Image analysis and machine learning for detecting malaria. [editorial flag: year and venue not stated in source document — verify before publication.]
- Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434*.
- Rivenson, Y., Ceylan Koydemir, H., Wang, H., et al. (2018). Deep learning enhanced mobile-phone microscopy. *ACS Photonics*, 5, 2354–2364.
- RSNA. (2020). *RSNA Pneumonia Detection Challenge*. Kaggle. Retrieved May 7, 2020, from <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
- Salehi, S., Abedi, A., Balakrishnan, S., & Gholamrezanezhad, A. (2020). Coronavirus disease 2019 (COVID-19) imaging reporting and data system (COVID-RADS) and common lexicon: A proposal based on the imaging data of 37 studies. <https://doi.org/10.1007/s00330-020-06863-0>
- Shi, F., Xia, L., Shan, F., Wu, D., Wei, Y., Yuan, H., Jiang, H., Gao, Y., Sui, H., & Shen, D. (2020). Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. [editorial flag: preprint/venue not fully specified in source — verify before publication.]
- Smith, Z. J., Chu, K., Espenson, A. R., et al. (2011). Cell-phone-based platform for biomedical device development and education applications. *PLoS ONE*, 6, e17150.
- Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2015). The emerging field of mobile health. *Science Translational Medicine*, 7, 283rv3.
- Syeda-Mahmood, T. (2018). Role of big data and machine learning in diagnostic decision support in radiology. *Journal of the American College of Radiology*.
- Thrall, J. H., Li, X., Li, Q., et al. (2018). Artificial intelligence and machine learning in radiology: Opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology*.
- Vashist, S. K., Mudanyali, O., Schneider, E. M., Zengerle, R., & Ozcan, A. (2014). Cellphone-based devices for bioanalytical sciences. *Analytical and Bioanalytical Chemistry*, 406, 3263–3277.
- West, D. (2012). How mobile devices are transforming healthcare. *Issues in Technology Innovation*, 18, 1–11.
- Wilson, M. L., Fleming, K. A., Kuti, M. A., et al. (2018). Access to pathology and laboratory medicine services: A crucial gap. *The Lancet*, 391(10133), 1927–1938.
- Wong, H. Y. F., Lam, H. Y. S., Fong, A. H.-T., et al. (n.d.). Frequency and distribution of chest radiographic findings in COVID-19 positive patients. [editorial flag: year and venue not stated in source document — verify before publication.]
- Yang, Z., & Zhan, Q. (2016). Single-shot smartphone-based quantitative phase imaging using a distorted grating. *PLoS ONE*, 11, e0159596.
- Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *arXiv:1809.07294*.
- Yue, W., Wang, Z., Chen, H., et al. (2018). Machine learning with applications in breast cancer diagnosis and prognosis. *Designs*, 2(2), 13.
- Zhang, J., Xie, Y., Li, Y., et al. (2020). COVID-19 screening on chest X-ray images using deep learning based anomaly detection. [editorial flag: preprint/venue not fully specified in source — verify before publication.]

APPENDICES / FIGURES REFERENCED IN THE SOURCE

The source document references figures (performance charts, a GAN schematic, and application screenshots) that were not transferred into this adaptation, as the underlying images were not supplied. These should be re-inserted from the original before publication:

- Fig. 1 — Iteration 6 performance metrics in Custom Vision
- Fig. 2 — Plain GAN schematic diagram
- Fig. 4 — Household objects data loaded in Custom Vision
- Fig. 6 — Smartphone app detecting COVID-19 X-ray and Inapplicable images
- Fig. 7 — Web application built with Custom Vision REST API
- Fig. 8 — Web application detecting COVID-19 with 100% confidence

Editorial Notes: Data-Completeness Checklist

- Confidence-interval table (Section 3.2) reports bounds outside the valid 0–100% range for a percentage metric — very likely a calculation or transcription artifact in the source. Recalculate and correct before this comparison is cited in a policy context.
- Figures 1, 2, 4, 6, 7, 8 referenced in the source were not supplied with this draft and need to be re-inserted from the original materials.
- Several references (flagged individually above) lack a publication year or venue in the source document and should be verified against the original bibliography before release.
- Third-party code/APK links (Android/iOS repositories, APK download) cited in Section 2.3 were not independently checked for this adaptation and should be verified as live before publication.
- Declaration of interests: the source states no conflicts of interest and no external funding; carried forward here as stated, not independently verified.

Suggested citation: Mustafiz, R., & Mohsin, K. (2026). *Assessing automated machine learning services to detect COVID-19 from X-ray and CT images: A real-time smartphone application case study* (CHRSD Foundation Research Paper Series adaptation). Centre for Humanitarian Research and Social Development. <https://chrdsd.org/research-papers>