Virus detection and identification in minutes using single-particle imaging and deep learning

Nicolas Shiaelis[1,*], Alexander Tometzki[1], Leon Peto[2,3], Andrew McMahon[1], Christof Hepp[1], Erica Bickerton[4], Cyril Favard[5], Delphine Muriaux[5,6], Monique Andersson[3], Sarah Oakley[3], Alison Vaughan[2,7], Philippa C. Matthews[2], Nicole Stoesser[2,8], Derrick Crook[2,7,8], Achillefs N. Kapanidis[1,*] and Nicole C. Robb[1,9,*]

[1]Biological Physics Research Group, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, OX1 3PU, United Kingdom

[2]Nuffield Department of Medicine, University of Oxford, Oxford, OX3 9DU, United Kingdom

[3]Department of Microbiology, Oxford University Hospitals NHS Foundation Trust, Oxford, OX3 9DU, United Kingdom

[4]The Pirbright Institute, Ash Road, Pirbright, Woking, Surrey, GU24 0NF, United Kingdom

[5]Membrane Domains and Viral Assembly, IRIM, UMR 9004 CNRS & University of Montpellier, 1919, route de Mende, 34293 Montpellier, France

[6]CEMIPAI, UMS 3725 CNRS & University of Montpellier, 1919, route de Mende, 34293 Montpellier, France

[7]NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, United Kingdom

[8]NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with Public Health England, University of Oxford, Oxford, United Kingdom

[9]Warwick Medical School, University of Warwick, Coventry, CV4 7AL, United Kingdom

*To whom correspondence should be addressed: Nicole.Robb@warwick.ac.uk, kapanidis@physics.ox.ac.uk, nicolas.shiaelis@st-annes.ox.ac.uk

Abstract

The increasing frequency and magnitude of viral outbreaks in recent decades, epitomized by the current COVID-19 pandemic, has resulted in an urgent need for rapid and sensitive viral diagnostic methods. Here, we present a methodology for virus detection and identification that uses a convolutional neural network to distinguish between microscopy images of single intact particles of different viruses. Our assay achieves labeling, imaging and virus identification in less than five minutes and does not require any lysis, purification or amplification steps. The trained neural network was able to differentiate SARS-CoV-2 from negative clinical samples, as well as from other common respiratory pathogens such as influenza and seasonal human coronaviruses, with high accuracy. Single-particle imaging combined with deep learning offers a promising alternative to traditional viral diagnostic methods, and has the potential for significant impact.

Main

The SARS-CoV-2 betacoronavirus has infected millions of people in 2020, resulting in over a million deaths, and causing worldwide social and economic disruption. Rapid, sensitive and accurate viral diagnosis is fundamental to response efforts.

Current SARS-CoV-2 diagnostic methods include nucleic acid amplification tests, antigen detection, and serology tests[1]. Reverse transcriptase polymerase chain reaction (RT-PCR) is considered the gold standard for diagnosis; however, RT-PCR takes several hours to provide a result, is restricted to specialized laboratories (as it requires viral lysis and RNA extraction), and can be limited by supply chain issues. Isothermal nucleic acid amplification methods, such as loop-mediated isothermal amplification (RT-LAMP), offer a promising alternative that does not require thermal cycling and can provide results within an hour[2-7]; however, these methods are still subject to similar supply chain issues as RT-PCR. Rapid (<30 minutes) immunoassay-based antigen-detecting tests exist for some viruses (e.g., influenza), but generally have low sensitivities[8]. There is thus an urgent need for new viral detection approaches, particularly ones that can be deployed in non-laboratory settings.

To address this need, we developed a novel diagnostic method that relies on the detection of intact virus particles using wide-field fluorescence imaging. Our method starts with near-instantaneous labeling of enveloped viruses via cation-mediated binding of short fluorescent DNAs to the surface of virus particles[9]; we subsequently surface-immobilise labelled particles, collect diffraction-limited images containing thousands of labelled particles, and finally use image analysis and machine-learning to identify different viruses in biological and clinical samples (Fig.1A). Our approach exploits the fact that distinct virus types and strains have differences in surface chemistry, size, and shape, which in turn affects the fluorophore distribution over the surface of different viruses. Such differences can be captured by convolutional neural networks (CNNs), which have been used previously to classify super-resolved microscopy images of heterogeneous virus populations into particle classes with distinct structural features[10], and to detect virus particles in transmission electron microscopy images[11].

To demonstrate our ability to label, immobilize, and image coronavirus particles, we initially used infectious bronchitis virus (IBV), an avian coronavirus (CoV). We labelled IBV using strontium chloride and a mixture of green and red fluorescent DNAs (labelled with Cy3 or Atto647N fluorophores, respectively); immobilized particles on a chitosan-coated glass slide; and imaged particles using total-internal-reflection fluorescence microscopy (Sup.Fig.1). Fluorescent labelling was achieved within seconds via a single-step addition of labelling mixture (see Methods), after which the viruses were immediately immobilized. The resulting images contained particles with either single green or red fluorescence signals (shown as green and red particles), as well as colocalised green and red fluorescence signals (shown as yellow particles) (Fig.1B,C). In a virus-negative control, substantially fewer colocalised signals were observed (Fig.1B,D), consistent with the fact that single red and green signals arose from free DNA, while the majority of colocalised signals corresponded to coronavirus particles labelled with both colors. Colocalised signals in the absence of virus likely occurred due to random coincidence or cation-mediated clustering of DNAs on the surface. Omission of DNAs resulted in complete loss of the fluorescent signal (Fig.1B, left panels).

Prior to use for machine learning, individual image signals were isolated into bounding boxes (BBXs) using segmentation of the field of view (FOV) through adaptive filtering. The raw FOVs from the red channel (Fig.1Ei) were converted into a binary format (Fig.1Eii) and area filtering used to disregard objects with an area smaller than 10 pixels (single fluorophores) or larger than 100 pixels (aggregates) (Fig.1Eiii). At the same time, to enrich our sampling for viruses and exclude signals not arising from virus particles, the

location image (showing the green, red and yellow signals from both channels; Fig.1Eiv) was used to identify colocalised signals (Fig.1Ev). This information was then combined with the signals identified in the filtered binary image (Fig.1Eiii) to reject signals not meeting the colocalisation condition (Fig.1Evi; cyan boxes) and retain signals meeting the colocalisation condition (Fig.1Evi-vii; red boxes). The segmentation was fully automated, allowing each FOV to be processed in ~2 seconds. The mean number of colocalised BBXs per FOV obtained when IBV was present was ~6-fold higher than when the virus was absent (Fig.1F).

Next, we tested whether our CNN architecture could differentiate between signals observed in virus-positive and virus-negative samples, as well as between images of different viruses. Many respiratory viruses, including SARS-CoV-2, influenza and seasonal human coronaviruses (hCoV), exhibit similar early onset symptoms; it is thus crucial that diagnostic assays can differentiate between these different viruses. As a proof-of-principle, we fluorescently labelled and imaged IBV and three laboratory-grown influenza A strains: H3N2 A/Udorn/72 (Udorn), H3N2 A/Aichi/68 (X31), and H1N1 A/PR8/8/34 (PR8) (Fig.2A). These viruses are similar in size and shape, and cannot be distinguished in diffraction-limited microscope images of fluorescently labelled particles (Sup.Fig.2A). After image segmentation (Sup.Fig.2B) and examination of the properties of the resulting BBXs, we observed that the four viruses exhibited small differences in maximum pixel intensity, area, and semi-major-to-semi-minor-axis-ratio within the BBXs (Fig.2B-D); e.g., IBV appears brighter than influenza, whereas Udorn occupies a larger area than the other viruses. Such features are not easily identifiable by manual analysis, however these and other image features such as pixel correlations, can be exploited by deep-learning algorithms for classification purposes[12,13].

To classify different viruses, we constructed a 15-layer CNN (Fig.2E, see legend for detail). Four independent datasets of each virus strain and a virus-negative control were randomly divided into a training dataset (70%) and a validation dataset (30%). The datasets used for both training and validation of the model consisted of data collected from three different days of experiments to enhance the ability of the trained models to classify data from future datasets. The network was trained using different combinations of all four viruses and the negative control, using ~3000 BBXs per sample. To validate our network, we checked if the network could differentiate IBV virus samples from negative controls consisting of only strontium chloride and DNA; the first data point in the network validation session was at 50% accuracy (as expected for a completely random classification of objects into two categories), followed by an initial rapid increase in validation accuracy as the network detected the most obvious parameters, followed by a slower increase as the number of iterations increased (Sup.Fig.3A). This was accompanied by a similar decrease in the Loss Function (Sup.Fig.3B); the entire training and validation process took 12 minutes to complete (Sup.Fig.3C).

Results of the network validation are shown as a confusion matrix, commonly used to visualize performance measures for classification problems (Fig.2F). The rows correspond to the predicted class (Output Class), the columns to the true class (Target Class), and the far-right, bottom cell represents the overall validation accuracy (hereafter, accuracy) of the model for each classified particle. The trained network was able to differentiate positive and negative IBV samples with high accuracy (91.4%), sensitivity (91.9%) and specificity (90.9%) (Fig.2G). Of note, these probabilities refer to <u>single virus particles</u> in the sample and *not* the whole sample; the probability of correctly identifying a sample with hundreds or thousands of IBV virus particles will therefore approach 100%.

Next, we tested the network's ability to distinguish between different virus types and strains by training the network on data from IBV and influenza Udorn, X31 and PR8 strains. The network easily distinguished between IBV and influenza, with an accuracy of 95.5% for IBV vs. Udorn (Fig.2H) and 94.3% for IBV vs. PR8 (Sup.Fig.4A). The network was able to differentiate between two strains of influenza (Udorn and X31),

albeit with a lower accuracy (68.8%), reflecting the greater homogeneity between strains of the same virus (Sup.Fig.4B). The network was also able to distinguish between IBV and a pooled dataset consisting of the virus-negative control and three influenza strains (92.2%) (Sup.Fig.4C).

Having validated our assay on laboratory-grown viruses, we next assessed clinical samples. Throat swabs from patients negative for virus, or positive for SARS-CoV-2, seasonal hCoVs (OC43, HKU1 or NL63) or human influenza A (as determined by RT-PCR) were inactivated with formaldehyde before being labelled and immobilised (see Methods). Images were captured on four different days, with data from days 1-3 used to train and validate the network (Table1, Sup.Fig.5). The trained network was able to distinguish between virus-positive and virus-negative samples with excellent accuracy, distinguishing between SARS-CoV-2-positive and negative BBXs with an accuracy of ~70% (Fig.3A), between Flu A and negative BBXs with an accuracy of ~87% (Fig.3B), and between seasonal hCoV and negative samples with an accuracy of ~78% (Sup.Fig.6A). The decrease in accuracy (compared to the laboratory-grown viruses) reflected the greater heterogeneity and complexity of clinical samples (e.g., varied storage conditions, wide range of virus concentrations, presence of residual cellular material, different sampling techniques). In spite of these issues, the network could also distinguish SARS-CoV-2 from seasonal hCoVs with a validation accuracy of ~73% (Fig.3C), and SARS-CoV-2 from Flu A with a validation accuracy of ~70% (Sup.Fig.6B), potentially useful in diagnosing co-circulating infections. Having trained and validated the network, data acquired on day 4 were then used to test the ability of the CNN to categorise the same samples imaged on a different day (Sup.Table1). The network was able to classify more than 50% of BBXs correctly in 8 of 10 samples tested for seasonal hCoVs vs. negative, and in 8 of 9 samples tested for SARS-CoV-2 vs. hCoV; results can be further improved by increasing the number of samples used for training.

We then tested our network's ability to diagnose independent clinical samples never seen before by the trained network; blind positive or negative samples were imaged on day 4 and assessed by the trained network within a few seconds. The output number of BBXs classified as positive or negative for each sample, and their associated probability values, were compared to the cumulative probability distribution functions (PDFs) expected for either positive or negative samples. When a SARS-CoV-2 RT-PCR-negative sample was analysed by the trained network, the large majority of BBXs were classified as negative (125 vs 29), and the associated probability of the sample being negative was overwhelmingly higher than the probability of it being positive (0.97 vs 1.64×10^{-63}) (Fig.3D), as expected. Similarly, when a SARS-CoV-2 RT-PCR-positive sample was analysed, the large majority of BBXs were classified as positive (148 vs 25) and the associated probability of the sample being positive was overwhelmingly higher than the probability of it being negative (1.00 vs 2.49×10^{-52}) (Fig.3E). Similar results were also obtained for an hCoV OC43 clinical sample (Sup.Fig.7).

We also estimated the limit of detection (LOD) of our assay by testing the ability of the network to accurately detect increasing IBV concentrations (Sup.Fig.8). Images were analysed by the trained network, and the number of particles classified as positive was fitted linearly with increasing virus concentration. The LOD was estimated as 6×10^4 PFU/mL, a sensitivity that, as expected, was significantly lower than that of amplification-based methods like RT-PCR (~10^2 PFU/mL[14]); however, we anticipate that the sensitivity should increase substantially with better immobilisation, pre-concentration of virus particles, optimised labelling, and reduced surface binding of free DNA.

Our work demonstrates how single-particle fluorescence microscopy combined with deep learning can help to rapidly detect and classify viruses, including coronaviruses. Our approach of instantaneous labelling, rapid automated imaging, pre-processing and deep learning classifies viruses within minutes, avoiding the need for viral lysis or amplification and the associated cost, tedium and supply-chain issues.

The non-specific detection of intact viral particles (rather than genome fragments) can report directly on infectivity, and has the advantages of speed (results within 2-5 minutes), the ability to detect multiple virus types in a single labelling step, and robustness against potential mutations in the viral genome. Our algorithms are extremely versatile and can be trained to differentiate between many different viruses, independently of how they are labelled, immobilized and imaged. Given its simplicity and rapid nature, our technology could also be used outside of specialized laboratories, such as in airports, workplaces and care homes. These unique capabilities should enable extremely rapid, mobile, and real-time analysis of patient and community samples during pandemic situations.

Methods

Laboratory grown virus strains and DNAs. The influenza strains (H1N1 A/Puerto Rico/8/1934 (PR8), H3N2 A/Udorn/72 (Udorn) and H3N2 A/Aichi/68 (X31)) used in this study have been described previously[9]. Briefly, PR8 and Udorn were grown in Madin-Darby bovine kidney (MDBK) or Madin-Darby canine kidney (MDCK) cells and X31 was grown in embryonated chicken eggs. The cell culture supernatant or allantoic fluid was collected and the viruses were titred by plaque assay. Titres of PR8, Udorn and X31 were 1.05×10^8 plaque forming units (PFU)/mL, 1.0×10^7 PFU/ mL and 4.5×10^8 PFU/mL respectively. The coronavirus IBV (Beau-R strain)[15] was grown in embryonated chicken eggs and titred by plaque assay (1×10^6 PFU/mL). Viruses were inactivated by addition of 2% formaldehyde before use.

Single-stranded oligonucleotides labelled with either red or green dyes were purchased from IBA (Germany). The 'red' DNA was modified at the 5' end with ATTO647N (5' ACAGCACCACAGACCACCCGCGGATGCCGGTCCCTACGCGTCGCTGTCACGCTGGCTGTTTGTCTTCCTGCC 3') and the 'green' DNA was modified at the 3' end with Cy3 (5' GGGTTTGGGTTGGGTTGGGTTTTTGGGTTTGGGTTGGGTTGGGAAAAA 3').

Clinical samples. Ethical approval was obtained for the use of anonymised oro- or nasopharyngeal specimens from patients for the diagnosis of influenza and other respiratory pathogens, including SARS-CoV-2 (North West-Greater Manchester South Research Ethics Committee [REC], REC Ref:19/NW/0730). Specimens were maintained in Copan Universal Transport Medium (UTM) before being inactivated in a 4% final concentration of formaldehyde (Pierce) for 30 minutes at room temperature. Samples were confirmed as SARS-CoV-2-positive or negative using either the Public Health England 2019-nCoV real-time RT-PCR RdRp gene assay or RealStar SARS-CoV-2 RT-PCR Kit (Altona diagnostics). Testing for other respiratory pathogens and sub-typing of seasonal human coronavirus (hCoV) samples as OC43, HKU1 or NL63 strains was conducted using the BioFire FilmArray Respiratory Panel (Biomerieux, Marcy-L'Etoile, France) and Cepheid Xpert Xpress Flu/RSV (Cepheid, Sunnyvale, CA, USA).

Sample preparation. Glass slides were treated with 0.015 mg/mL chitosan (a linear polysaccharide) in 0.1 M acetic acid for 30 min before being washed thrice with MilliQ water. Unless otherwise stated, virus stocks (typically 10 µL) were diluted in 0.23 M $CaCl_2$ or $SrCl_2$ (as described in the figure legends) and 1 nM of each fluorescently-labelled DNA in a final volume of 20 µL, before being added to the slide surface. Virus labelling with $CaCl_2$ has been described previously[9]; $SrCl_2$ provides similar results. For laboratory grown virus stocks, negatives were taken using Minimal Essential Media (Gibco) or allantoic fluid from uninfected eggs in place of the virus.

Imaging. Images were captured using a wide-field fluorescence microscope, as previously described[9]. The sample was imaged using total internal reflection fluorescence (TIRF) microscopy. The laser illumination was focused at a typical angle of 53° with respect to the normal. Movies of 5 frames per field of view (FOV) (measuring 75 x 49 µm) were taken at a frequency of 33 Hz and exposure time of 30 ms, with laser

intensities kept constant at 0.78 kW/cm^2 for the red (640 nm) and 1.09 kW/cm^2 for the green (532 nm) laser. To automate the task and ensure no bias in the selection of FOVs, the whole sample was scanned using the multiple acquisition capability of the microscope; 81 FOVs were imaged in 2 minutes.

Data Segmentation. Each FOV in the red channel was turned into a binary image using MATLAB's built-in imbinarize function with adaptive filtering sensitivity set to 0.5. Adaptive filtering uses statistics about the neighbourhood of each pixel it operates on to determine whether the pixel is foreground or background. The filter sensitivity is a variable which, when increased, makes it easier to pass the foreground threshold. The bwpropfilt function was used to exclude objects with an area outside the range 10-100 pixels, aiming to disregard free ssDNA and aggregates. The regionprops function was employed to extract properties of each found object: area, semi-major to semi-minor axis ratio (or simply, axis ratio), coordinates of the object's centre, bounding box (BBX) encasing the object, and maximum pixel intensity within the BBX.

Accompanying each FOV is a location image (LI) summarising the locations of signals received from each channel (red and green); colocalised signals in the LI image were shown in yellow. Objects found in the red FOV were compared with their corresponding signal in the associated LI. Objects that did not arise from colocalised signals were rejected. The qualifying BBXs were then drawn onto the raw FOV and images of the encased individual viruses were saved.

Machine Learning. The bounding boxes (BBX) from the data segmentation had variable sizes, however due to the size filtering they were never larger than 17 pixels in any direction. Thus, all the BBX were augmented such that they had a final size of 17x17 pixels by means of padding (adding extra pixels with 0 grey-value until they reach the required size). The augmented images were fed into the 15-layer CNN. The network had 3 convolutional layers in total, with kernels of 2x2 for the first two convolutions and 3x3 for the last one. The learning rate was set to 0.01 and the learning schedule rate remained constant throughout the training. The hyperparameters remained the same throughout the training process for all models; the mini batch size was set to 1000, the maximum number of epochs to 100 and the validation frequency to 20.

In the classification layer, trainNetwork took the values from the softmax function and assigned each input to one of K mutually exclusive classes using the cross entropy function for a 1-of-K coding scheme[16],

$$loss = \sum_{i=1}^{N} \sum_{j=1}^{K} t_{ij} \ln(y_{ij}) \qquad (1)$$

where N is the number of samples, K is the number of classes, t_{ij} is the indicator that the ith sample belongs to the jth class, and y_{ij} is the output for sample i for class j, which in this case, is the value from the softmax function. That is, it is the probability that the network associates the ith input with class j[17]. A stochastic gradient descent with momentum set to 0.9 was used as the optimizer.

Statistical Analysis. The results of each network validation are shown as a confusion matrix, which make used of the following terms:

- True positive (TP): BBXs correctly identified as positive,
- False Positive (FP): BBXs incorrectly identified as positive,
- True negative (TN): BBXs correctly identified as negative, and
- False negative (FN): BBXs incorrectly identified as negative.

Sensitivity refers to the ability of the test to correctly identify positive BBXs. It can be calculated by dividing the number of true positives over the total number of positives[18].

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

Specificity refers to the ability of the test to correctly identify negative BBXs. It can be calculated by dividing the number of true negatives over the total number of negatives[18].

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

The percentages of BBXs that are correctly and incorrectly predicted by the trained model are known as the positive predictive value (PPV) and negative predictive value (NPV) respectively.

$$PPV = \frac{TP}{TP+} \quad (4)$$

$$NPV = \frac{TN}{FN+TN} \quad (5)$$

The overall balanced validation accuracy of the model is given by:

$$Balanced\ Accuracy = \frac{Sensitivity+Specificit}{2} \quad (6).$$

A description of the cumulative probability distribution function for the blind test is given by:

$$f(x|N,p) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{N-1} \quad ; \quad x = 0, 1, 2, \dots, N \quad (7)$$

Where x is the number of the positively classified BBXs in the total number N of BBXs of the sample under consideration, and p is the sensitivity of the trained network. The result is the probability that at most x number of BBXs are classified as positive/negative, out of a total number N of BBXs.

In order to calculate the limit of detection (LOD), increasing concentrations of CoV (IBV; dilutions in allantoic fluid) were labelled and imaged, the resulting images were pre-processed, and the individual signals were fed into the trained network. The normalised average of *TP (TP/TP+FP)* and standard deviation (STD) were calculated and plotted against the corresponding concentrations as a scatter plot. The plot was fitted as a linear regression, as given by:

$$y = ax + b \quad (8)$$

Where the virus concentration was treated as the independent variable and *a* represents the LOD. For the final value of the LOD *a+(3STD)=6*10^4 PFU/mL* was used, which corresponds to a 99.85% confidence interval assuming a normal distribution.

Acknowledgements

We are grateful to Micron Oxford, funded by Wellcome Strategic Awards (091911 and 107457; PI Ilan Davis), for their loan of their microscope and to Nadia Halidi for her help with the instrument. This research was supported by a Royal Society Dorothy Hodgkin Research Fellowship DKR00620 and Research Grant for Research Fellows RGF\R1\180054 (N.C.R.), the University of Oxford COVID-19 Research Response Fund (N.C.R and A.N.K.), a BBSRC-funded studentship (N.S.), and Wellcome Trust grant 110164/Z/15/Z (A.N.K.). All data will be available upon request.

Author Contributions

N.S., A.T., A.M., C.H., A.N.K. and N.C.R designed and carried out experiments, analysed data and interpreted results. N.S. wrote analysis software. E.B. provided reagents. C.F. and D.M. validated the virus inactivation method. L.P., M.A., S.O., A.V., P.C.M., N.S. and D.C. collected, diagnosed and typed clinical

samples that were considered to be appropriate for the purposes of this study. N.C.R. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing Interests statement

This work was supported by the National Institute for Health Research (NIHR) Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at University of Oxford in partnership with Public Health England (PHE), and the NIHR Oxford Biomedical Research Centre. The report presents independent research. The views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR, or the Department of Health.

The work was carried out using portable commercial microscopes from Oxford Nanoimaging, a company in which A.N.K. is a co-founder and shareholder. Patent applications relating to the work have been submitted by N.C.R., A.N.K. and N.S. (PCT/GB2019/053073 and 2006155.6).

References

1 Udugama, B. *et al.* Diagnosing COVID-19: The Disease and Tools for Detection. *ACS Nano* **14**, 3822-3835, doi:10.1021/acsnano.0c02624 (2020).

2 Huang, W. E. *et al.* RT-LAMP for rapid diagnosis of coronavirus SARS-CoV-2. *Microb Biotechnol* **13**, 950-961, doi:10.1111/1751-7915.13586 (2020).

3 Lu, R. *et al.* Development of a Novel Reverse Transcription Loop-Mediated Isothermal Amplification Method for Rapid Detection of SARS-CoV-2. *Virol Sin*, doi:10.1007/s12250-020-00218-1 (2020).

4 El-Tholoth, M., Bau, H. H. & Song, J. A Single and Two-Stage, Closed-Tube, Molecular Test for the 2019 Novel Coronavirus (COVID-19) at Home, Clinic, and Points of Entry. *ChemRxiv*, doi:10.26434/chemrxiv.11860137 (2020).

5 Lamb, L., Bartolone, SN, Elijah, W, Chancellor, MB. Rapid Detection of Novel Coronavirus (COVID-19) by Reverse Transcription-Loop-Mediated Isothermal Amplification. *medRxiv*, doi:https://doi.org/10.1101/2020.02.19.20025155 (2020).

6 Zhang, Y., Odiwuor, N., Xiong, J., Sun, L., Nyaruaba, RO, Wei, H, Tanner, NA. Rapid Molecular Detection of SARS-CoV-2 (COVID-19) Virus RNA Using Colorimetric LAMP. *medRxiv*, doi:https://doi.org/10.1101/2020.02.26.20028373 (2020).

7 Rodriguez-Manzano, J., Malpartida-Cardenas, K., Moser, N., Pennisi, I., Cavuto, M., Miglietta, L., Moniri, A., Penn, R., Satta, G., Randell, P., Davies, F., Bolt, F., Barclay, W., Holmes, A., Georgiou, P. A handheld point-of-care system for rapid detection of SARS-CoV-2 in under 20 minutes. *medRxiv* (2020).

8 Chartrand, C. & Pai, M. How accurate are rapid influenza diagnostic tests? *Expert Rev Anti Infect Ther* **10**, 615-617, doi:10.1586/eri.12.49 (2012).

9 Robb, N. C. *et al.* Rapid functionalisation and detection of viruses via a novel Ca(2+)-mediated virus-DNA interaction. *Sci Rep* **9**, 16219, doi:10.1038/s41598-019-52759-5 (2019).

10 Laine, R. F. *et al.* Structured illumination microscopy combined with machine learning enables the high throughput analysis and classification of virus structure. *Elife* **7**, doi:10.7554/eLife.40183 (2018).

11 Ito, E., Sato, T., Sano, D., Utagawa, E. & Kato, T. Virus Particle Detection by Convolutional Neural Network in Transmission Electron Microscopy Images. *Food Environ Virol* **10**, 201-208, doi:10.1007/s12560-018-9335-7 (2018).

12 Rawat, W. & Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput* **29**, 2352-2449, doi:10.1162/NECO_a_00990 (2017).

13 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444, doi:10.1038/nature14539 (2015).
14 Arnaout, R. *et al.* SARS-CoV2 Testing: The Limit of Detection Matters. *bioRxiv*, doi:10.1101/2020.06.02.131144 (2020).
15 Casais, R., Thiel, V., Siddell, S. G., Cavanagh, D. & Britton, P. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol* **75**, 12359-12369, doi:10.1128/JVI.75.24.12359-12369.2001 (2001).
16 Bishop, C. M. *Pattern Recognition and Machine Learning*. (Springer, New York, NY, 2006).
17 Mathworks. <https://uk.mathworks.com/help/deeplearning/ug/layers-of-a-convolutional-neural-network.htmlbvk87b1> (2019).
18 Altman, D. G. & Bland, J. M. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* **308**, 1552, doi:10.1136/bmj.308.6943.1552 (1994).

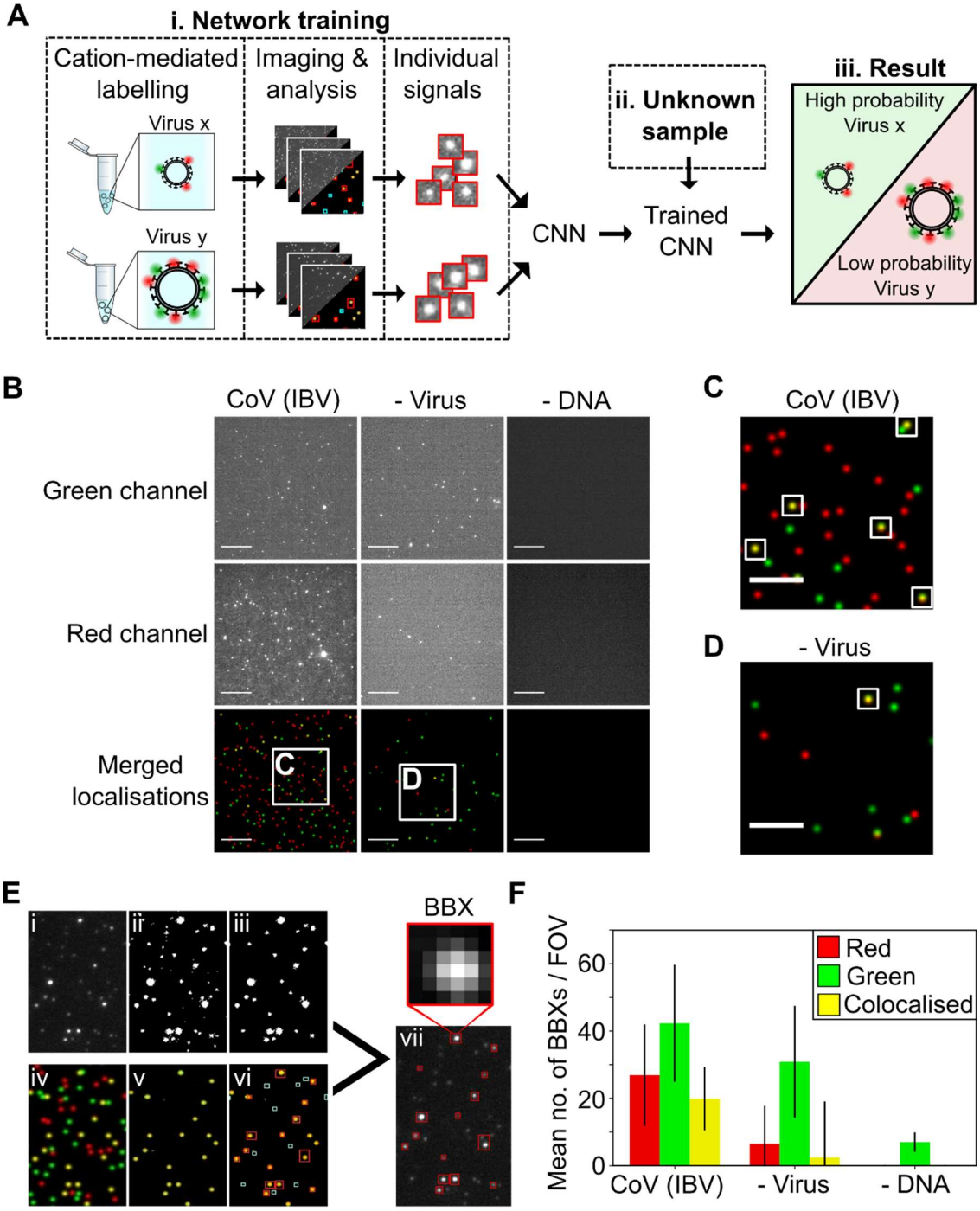

Figure 1. Design and validation of a fluorescent labelling strategy for coronavirus.
A) Overview of the diagnostic test. i) Viruses were fluorescently labelled and diffraction-limited images of them were acquired. Image analysis was used to isolate individual signals in the images and a deep-learning convolutional neural network (CNN) was trained to exploit differences in the features of different viruses to identify them. ii) Individual signals from images of unknown samples can then be fed into the trained CNN to allow iii) classification of virus particles.

B) Representative fields of view (FOVs) of fluorescently labelled infectious bronchitis virus (CoV (IBV)). The virus sample, at a final concentration of $1x10^4$ PFU/mL, was immobilized and labelled with 0.23M $SrCl_2$, 1nM Cy3 (green) DNA and 1nM Atto647N (red) DNA before being imaged. Green DNA was observed in the green channel (top panels) and red DNA in the red channel (middle panels); merged red and green localisations are shown in the lower panels. Scale bar 10µm. A negative control where DNA was replaced with water is included.

C&D) Zoomed in images from B), white boxes represent examples of colocalised particles. Scale bar 5µm.

E) Schematic of the segmentation process. (i) A single raw FOV from the red channel (cropped for magnification). (ii) Intensity filtering applied to i) to produce a binary image. (iii) Area filtering applied to ii) to include only the objects with areas between 10-100 pixels, thus excluding free DNA and aggregates. (iv) The location image associated with i). (v) Colocalised signals in the location image. (vi) Bounding boxes (BBXs) found from iii) drawn onto v). Objects that do not meet the colocalisation condition (shown in cyan) are rejected. (vii) BBXs of objects that do meet the colocalisation condition (shown in red) are drawn over i). Scale bar 10µm.

F) Plot showing the mean number of BBXs per FOV for labelled CoV (IBV) and the negative controls. Error bars represent the standard deviation from multiple FOVs.

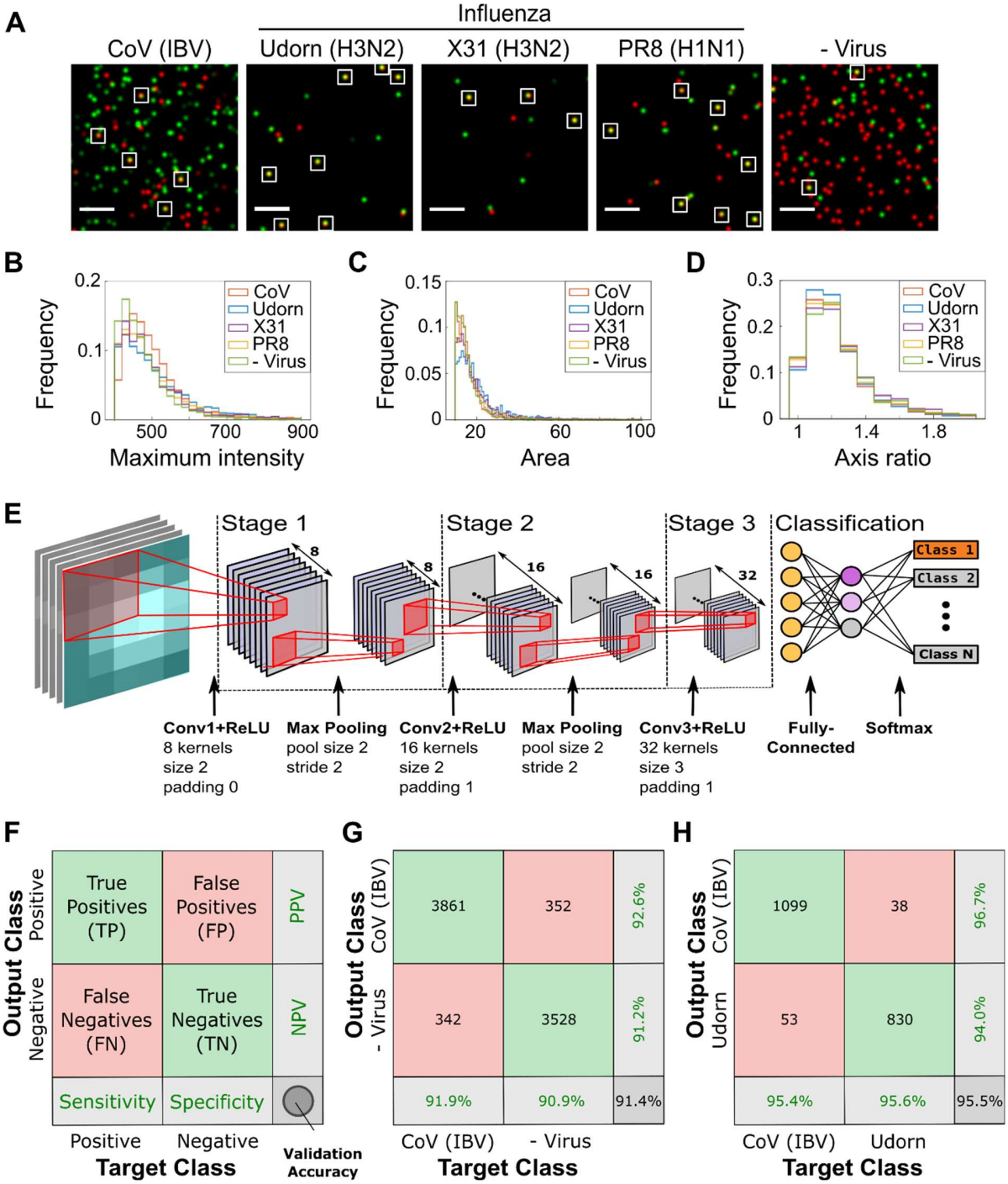

Figure 2. Design of a convolutional neural network to classify viruses.
A) Representative FOVs of fluorescently labelled coronavirus (CoV (IBV)), two strains of H3N2 influenza (A/Udorn/72 (Udorn) and A/Aichi/68 (X31)), an H1N1 influenza strain (A/PR8/8/34 (PR8)) and a negative control where virus was substituted with allantoic fluid. The samples were immobilized and labelled with 0.23M $SrCl_2$, 1nM Cy3 (green) DNA and 1nM Atto647N (red) DNA before being imaged. Merged red and

green localisations are shown, examples of colocalisations are highlighted with white boxes. Scale bar 10µm.

B-D) Normalised frequency plots of the maximum pixel intensity, area and semi-major-to-semi-minor-axis-ratio within the BBXs of the four different viruses.

E) Illustration of the 15-layer shallow convolutional neural network. Following the input layer (inputs comprising BBXs from the segmentation process), the network consists of three convolutions (stages 1-3). Stages 1 and 2 each contain a ReLU layer to introduce non-linearity, a batch normalisation layer (not shown) and a max pooling layer, while stage 3 lacks a max-pooling layer. The classification stage has a fully-connected layer and a softmax layer to convert the output of the previous layer to a normalised probability distribution, allowing the initial input to be classified.

F) Network validation results are shown as a confusion matrix; rows correspond to the predicted class (output class) and columns to the true class (target class), diagonal entries represent the percentage of correctly classified viruses and off-diagonal elements the false positives and negatives. The far-right column corresponds to the percentages of BBXs that are correctly and incorrectly predicted by the trained model, known as the positive and negative predictive values respectively, and the bottom row shows the percentages of all the BBXs belonging to each class that are correctly and incorrectly classified, also known as the sensitivity and specificity. The far-right, bottom cell represents the overall balanced validation accuracy of the model.

G) Confusion matrix showing that the trained network could differentiate positive and negative CoV (IBV) samples with high confidence. H) Confusion matrix showing that the trained network could differentiate between CoV (IBV) and influenza (Udorn).

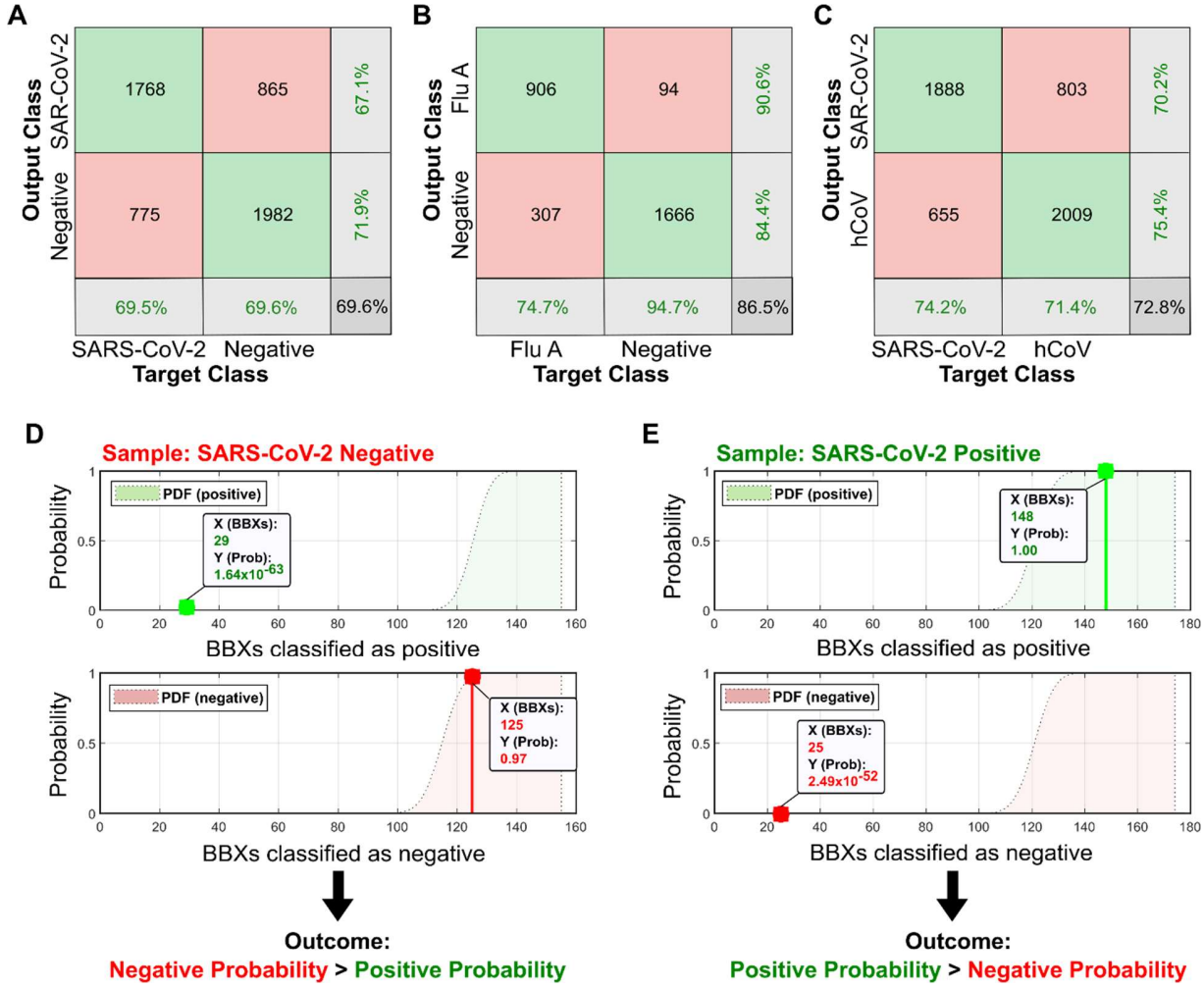

Figure 3. A deep learning network can diagnose SARS-CoV-2 in clinical samples.
A) Confusion matrix showing that a trained network could differentiate positive and negative SARS-CoV-2 clinical samples.

B) Confusion matrix showing that a trained network could differentiate influenza A (Flu A) positive clinical samples from negative samples.

C) Confusion matrix showing that a trained network could differentiate SARS-CoV-2 samples from seasonal human coronavirus (hCoV) samples.

D) Output from running a blind SARS-CoV-2 RT-PCR-negative sample through the trained network. The cumulative probability distribution function (PDF) assuming a sample is positive is shown in green (top panel), or assuming a sample is negative is shown in red (lower panel). Number of BBXs (X) and their associated probability values (Y) are given.

E) Output from running a blind SARS-CoV-2 RT-PCR-positive sample through the trained network.

				Number of BBXs				
	Virus type	RT-PCR	CT Value	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Total
1	SARS-CoV-2	POS	15.1	7954	2172	13915	0	24041
2	SARS-CoV-2	POS	15.4	370	223	131	0	724
3	SARS-CoV-2	POS	16.5	662	214	1340	0	2216
4	SARS-CoV-2	POS	12.3	861	218	1783	0	2862
5	SARS-CoV-2	POS	17.0	0	0	0	173	173
							Total	30016
7	HKU1	POS	---	229	104	203	0	536
8	HKU1	POS	---	169	92	136	0	397
9	HKU1	POS	---	330	129	148	0	607
							Total	1540
11	NL63	POS	---	377	80	279	0	736
12	NL63	POS	---	2488	513	823	0	3824
13	NL63	POS	---	453	207	292	0	952
							Total	5512
14	OC43	POS	---	268	131	254	0	653
15	OC43	POS	---	897	569	342	0	1808
16	OC43	POS	---	8805	145	0	0	8950
17	OC43	POS	---	0	0	0	2239	2239
							Total	13650
18	---	NEG	---	1507	337	4836	0	6680
19	---	NEG	---	562	77	45	0	684
20	---	NEG	---	451	568	150	0	1169
21	---	NEG	---	378	158	571	0	1107
22	---	NEG	---	0	0	0	319	319
23	---	NEG	---	0	0	0	354	354
							Total	10313

Table 1. Data on clinical samples used in the study. POS = positive, NEG = negative. Orange rows refer to samples used for blind testing. HKU1, NL63 and OC43 refer to samples positive for seasonal human coronaviruses (hCoV).