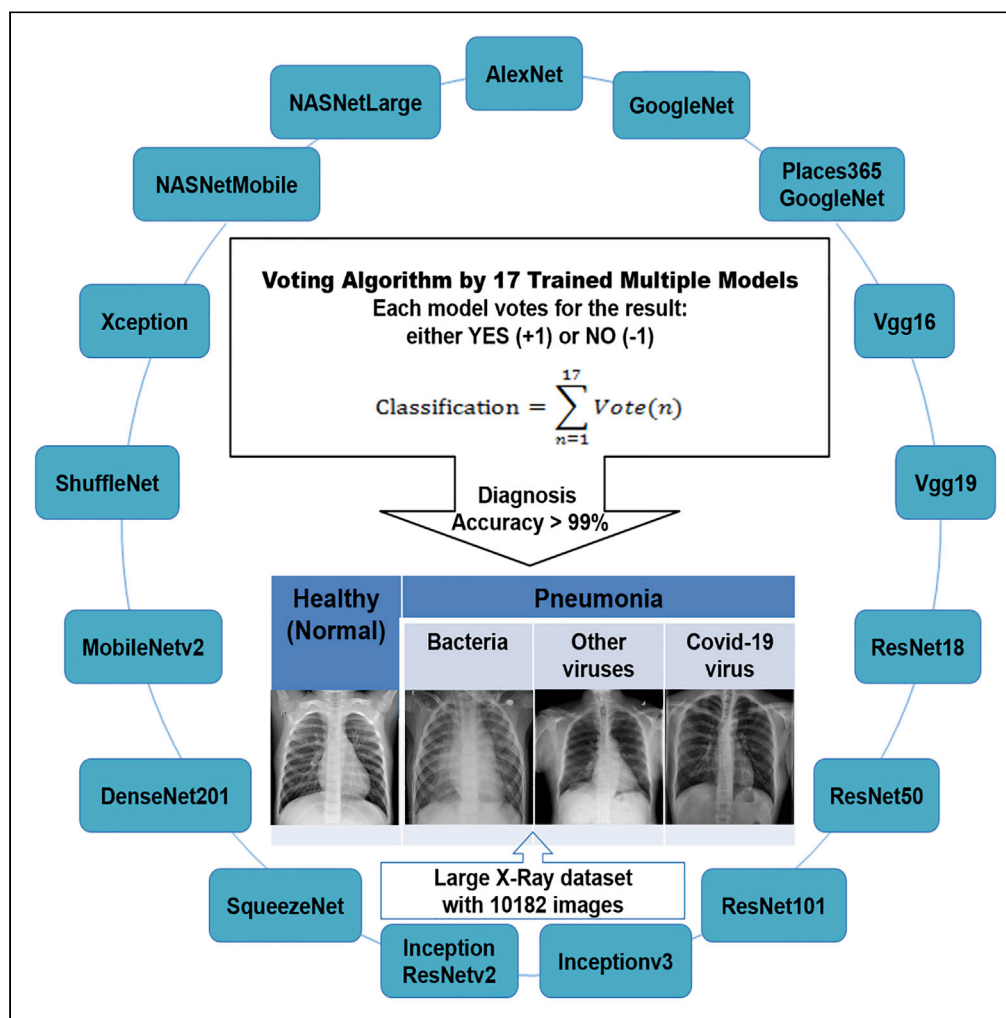


Article

An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images



Dongguang Li,
Shaoguang Li

shaoguang.li@umassmed.edu

Highlights

We used artificial intelligence models to diagnose Covid-19 pneumonia by reading chest X-ray images

We employed our unique deep learning voting algorithms combining multiple Convolutional neural networks

Our AI models reached a high diagnostic accuracy (>99%) for Covid-19 pneumonia detection

We obtained and analyzed a large chest X-ray image dataset (10,182 images)

Article

An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images

Dongguang Li¹ and Shaoguang Li^{1,2,*}**SUMMARY**

The coronavirus disease of 2019 (Covid-19) causes deadly lung infections (pneumonia). Accurate clinical diagnosis of Covid-19 is essential for guiding treatment. Covid-19 RNA test does not reflect clinical features and severity of the disease. Pneumonia in Covid-19 patients could be caused by non-Covid-19 organisms and distinguishing Covid-19 pneumonia from non-Covid-19 pneumonia is critical. Chest X-ray detects pneumonia, but a high diagnostic accuracy is difficult to achieve. We develop an artificial intelligence-based (AI) deep learning method with a high diagnostic accuracy for Covid-19 pneumonia. We analyzed 10,182 chest X-ray images of healthy individuals, bacterial pneumonia, and viral pneumonia (Covid-19 and non-Covid-19) to build and test AI models. Among viral pneumonia, diagnostic accuracy for Covid-19 reaches 99.95%. High diagnostic accuracy is also achieved for distinguishing Covid-19 pneumonia from bacterial pneumonia (99.85% accuracy) or normal lung images (100% accuracy). Our AI models are accurate for clinical diagnosis of Covid-19 pneumonia by reading solely chest X-ray images.

INTRODUCTION

Only within the first year of the pandemic, more than 100 million people worldwide have been diagnosed with Covid-19 that is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Lam et al., 2020; Ziegler et al., 2020) and proximate 2% death rate has been observed. In addition, fast-spreading SARS-CoV-2 variants have been identified worldwide (Grubaugh et al., 2020; Hu et al., 2020; Kirby, 2021; Korber et al., 2020; Tang et al., 2020). Although the majority of Covid-19 patients have mild symptoms and do not need specific treatment, approximately 15% of the patients end up developing severe pneumonia that potentially progresses to acute respiratory distress syndrome (ARDS) (Huang et al., 2020; Xu et al., 2020), which requires immediate intensive care to save lives. Pneumonia is often caused by viruses (including Covid-19) and bacteria and treatment options vary with the causes. Development of ARDS is a major cause of death in Covid-19 patients because of severe damage to the lungs (Wang et al., 2020; Zhang et al., 2020a). A correct clinical diagnosis of Covid-19 pneumonia guides appropriate treatment, especially with anti-SARS-CoV-2 immunotherapy. The SARS-CoV-2 virus can be detected by an RNA test, but this test does not reflect clinical features and severity of Covid-19; in addition, false negative and false positive cases have also been observed with RNA testing for SARS-CoV-2 virus detection (Gao et al., 2020; Zhang et al., 2020b). In addition, pneumonia in Covid-19-positive patients could be caused by non-Covid-19 viruses or bacteria, especially in a flu season, and needs to be distinguished from pneumonia caused by non-Covid-19 organisms. Thus, a correct clinical diagnosis of Covid-19 provides a solid base for initiating Covid-19-specific therapies and patient isolation. Chest X-ray and CT images are frequently taken for determining the presence of Covid-19 pneumonia by radiologists, but high diagnostic accuracy is difficult to achieve. On the other hand, the CT examination is often limited to larger hospitals. By contrast, chest X-ray is widely available, economical, and less time-consuming. Furthermore, chest X-ray helps to assess the severity and therapy-response to Covid-19 (Cohen et al., 2020; Hussain et al., 2020; Kikisetti et al., 2020; Shen et al., 2021; Wong et al., 2021; Zhu et al., 2020). However, the current challenge in evaluating chest X-ray images in Covid-19 patients is to achieve high diagnostic accuracy.

Deep learning is a type of machine learning in which a model learns to perform classification tasks directly from images. Deep learning is usually implemented using neural network architecture

¹Division of Hematology/Oncology, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA

²Lead contact

*Correspondence: shaoguang.li@umassmed.edu

<https://doi.org/10.1016/j.isci.2022.104031>



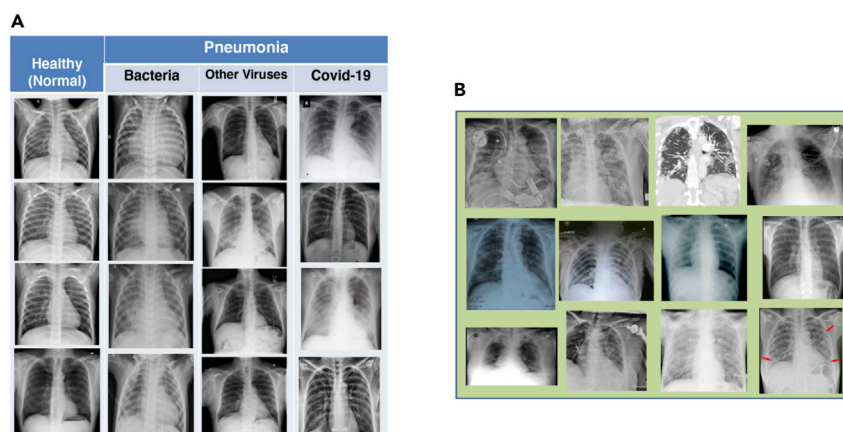


Figure 1. Chest X-ray image variations derived from sample collection

(A) There were some major variations of collected chest X-ray images in darkness, contrast, size, orientation, etc.

(B) There were some nonhuman structures in the chest X-ray images, including image labeling, pen marks, pictures of medical treatment devices, etc.

(Schmidhuber, 2015). Transfer learning is an approach that applies knowledge of one type of problem to a different but related problem (Weiss et al., 2016). Using a pretrained network with transfer learning is typically much faster and easier than training a network from scratch. Medical image analysis and computer-assisted intervention problems have been increasingly addressed with deep-learning-based solutions (Campanella et al., 2019). Although the available deep learning platforms are flexible, they do not provide specific functionality for medical image analysis and their adaption for this domain of application requires substantial implementation effort (Razavian, 2019). Consequently, there has been substantial duplication of effort and incompatible infrastructure has been developed across many research groups (Appenzeller, 2017). Furthermore, deep learning helps to generate computational models consisting of multiple processing layers to learn representations of data with multiple levels of abstraction (LeCun et al., 2015). Compared to machine learning that learns to conduct classification tasks directly from data, deep learning learns and abstracts the relevant information automatically while the data are being processed. Within the deep learning networks, the processing layers are interconnected via nodes (neurons), and each hidden layer receives information from the previous layer. Of all deep learning networks, convolutional neural networks (CNNs) are most commonly used, because CNNs can transform a multi-dimensional input image into a desired output (LeCun and Bengio, 1998). In general, a CNN is composed of an input layer and an output layer with several hidden layers in between and the most common layers are convolution, activation or ReLU (rectified linear activation function unit), and pooling. Each layer learns to detect different features in the input data. Deep learning networks have been widely used in the artificial intelligence (AI) field for signal data classification and we believe that they could be powerful tools for analyzing X-ray images.

Using deep learning, here we generate AI models to only read chest X-ray images of patients to reach nearly 100% diagnostic accuracy for Covid-19.

RESULTS

Chest X-ray image variations unrelated to pneumonia

To develop accurate AI models for reading chest X-ray images, high quality images are required as the quality of the images affects the outcomes. We found that sample variations were huge, posing a difficult challenge in building accurate AI models for reading chest X-ray images. A major type of sample variation was related to image collection, reflected by image darkness, contrast, size, orientation, (Figure 1A) etc. In addition, some images contained unexpected nonhuman structures such as image labeling, pen marks, pictures of medical treatment devices, (Figure 1B) etc. We realized that we need to build our AI models capable of identifying and excluding those non-pneumonia variations when reading chest X-ray images. It is equally important that our AI models should have abilities to distinguish Covid-19 pneumonia from non-Covid-19 pneumonia caused by viruses or bacteria.

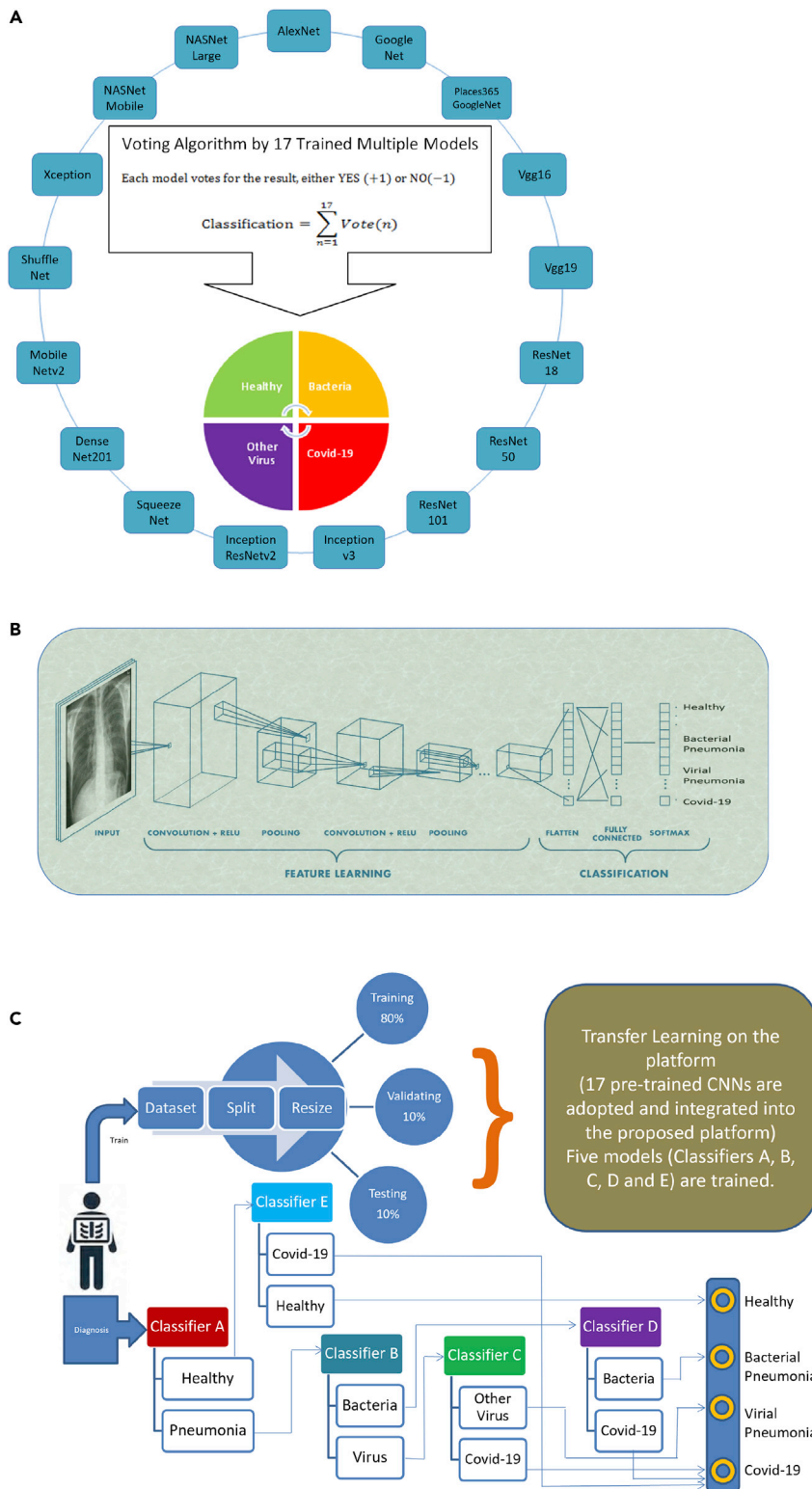


Figure 2. Strategies for overcoming image variations using AI models

(A) Core voting algorithm. The voting algorithm was developed by combining 17 trained AI models. Each model votes for either of the following results: Yes (+) or No (–).

Figure 2. Continued

(B) Architecture of a CNN comprising commonly used layers such as convolution, ReLU, and pooling.

(C) Overview of training and diagnosis. X-ray images were divided into classification groups based on the causes of pneumonia: Covid-19 virus, other viruses, bacteria, and healthy (as a control). 80% of the X-ray images were used for model training and 10% of the images for model validation, with the remaining 10% of the images for model testing.

Strategies for overcoming image variations using AI models

As described above, we faced a difficult challenge to overcome huge variations of chest X-ray images when generating highly accurate AI models for the diagnosis of Covid-19 pneumonia. It was clear to us that employment of a single deep CNN, as other researchers often did in establishing AI models, would fail to generate an AI model that allows achieving a high diagnostic accuracy upon reading chest X-ray images of patients. Therefore, we developed a unique voting algorithm that allowed for combining 17 CNNs and utilizing them as a whole to generate our AI models for optimizing the fitness of the data (Figure 2A), because we had a success in employing our AI models to read pathologic tissue images of patients with diffuse large B-cell lymphoma with a 100% diagnostic accuracy (Li et al., 2020a). The architecture of the CNN used in this study comprised multiple layers including convolution, ReLU, and pooling (Figure 2B). This advanced deep learning neural network approach was aimed to predict Covid-19 disease in patients and increase diagnostic accuracy using classification models with multiple CNNs based on deep learning (Figure 2B). We believe that the use of our voting algorithm with the combined 17 CNNs would be required for reconciling huge image variations to achieve high diagnostic accuracy in the clinic.

We totally reviewed 10,182 chest X-ray images obtained from public datasets for healthy individuals (5,510 cases), bacterial pneumonia (2,530), and viral pneumonia (non-Covid-19: 1,345 cases; Covid-19: 797 cases). We divided those X-ray images into classification groups based on the causes of pneumonia: Covid-19 virus, other viruses, bacteria, and healthy (as a control) (Figure 2C). We used 80% of the X-ray images for model training and 10% of the images for model validation, with the remaining 10% of the images for model testing. We expected that our combined 17 CNNs approach would help us to establish powerful AI models for reading chest X-ray images of Covid-19 pneumonia and distinguishing them from the images of other types of pneumonia with high diagnostic accuracy.

Achievement of high diagnostic accuracy with deep learning

We took a binary classification approach using deep learning; i.e., we set up each comparison group as Covid-19 versus non-Covid-19 (other viruses, bacteria, or healthy) with the combined 17 CNNs for classification (Figure 3A). On the other hand, we realized that training a deep CNN from scratch is computationally expensive and often requires a large amount of training data (about a few millions) which is not available in any public database including the one we used. Therefore, we used transfer learning (Figure 3A). We generated a particular AI model for each comparison group: classifier A (pneumonia vs healthy), classifier B (viruses vs bacteria), classifier C (Covid-19 vs other viruses), classifier D (Covid-19 vs bacteria), or classifier E (Covid-19 vs healthy) (Figure 3B). Specifically, we divided X-ray images into classification groups based on the causes of pneumonia: Covid-19 virus, other viruses, bacteria, and healthy (as a control). 80% of the X-ray images were used for model training and 10% of the images for model validation, with the remaining 10% of the images for model testing. Before performing the classification with multiple CNNs, each CNN was trained (or fine-tuned) with optimized parameters to achieve reasonably good performance. Those parameters include learning rate (0.001–0.0001), validation frequency (10–30), mini batch size (16–64), max, epochs (20–50), and algorithms (sgdm, adam, and rmsprop). In the end, all trained CNNs were fed into a platform where our core voting algorithm made them work together to produce the final classification results. With a focus on specifically identifying Covid-19 pneumonia and distinguishing it from pneumonia caused by other viruses (classifier C) and bacteria (classifier D), we found that our AI models achieved 99.95% diagnostic accuracy for Covid-19 from reading chest X-ray images of virus-caused pneumonia and achieved 99.85% accuracy for Covid-19 from reading the images of pneumonia caused by Covid-19 and bacteria (Figure 3C). We also generated an AI model for reading the Covid-19 and healthy images (classifier E) and the diagnostic accuracy for Covid-19 reached 100% (Figure 3C). These results demonstrate that our AI models provide accurate diagnosis of Covid-19 through reading chest X-ray images for clinical use.

In a practical sense, a patient with suspected lung infection demands further examination for pneumonia, for example, by X-ray. If pneumonia existed, it is beneficial to determine whether pneumonia is caused by

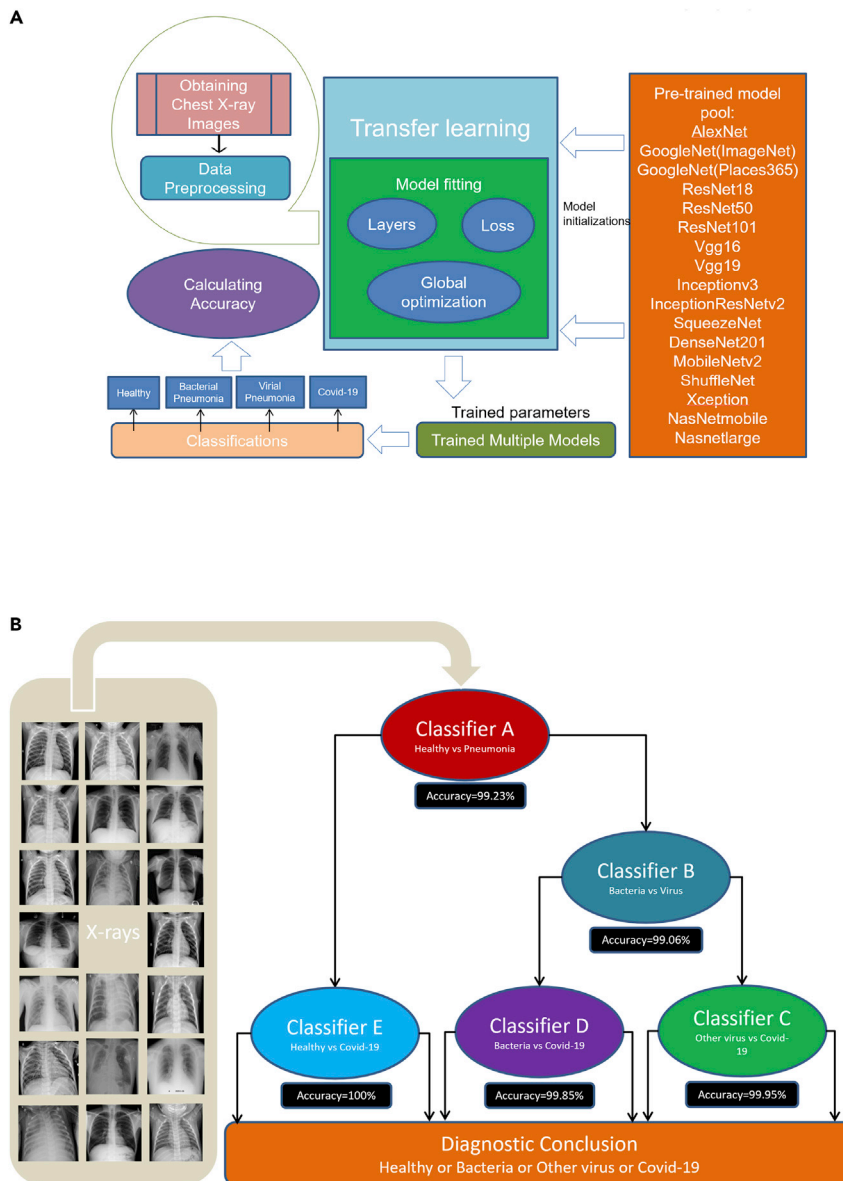


Figure 3. Achievement of high diagnostic accuracy with deep learning

(A) A binary classification approach using deep learning. Each comparison group was set up as Covid-19 versus non-Covid-19 (other viruses, bacteria, or healthy) with the combined 17 CNNs for classification. A particular AI model was generated for each comparison group: classifier A (pneumonia vs healthy), classifier B (viruses vs bacteria), classifier C (Covid-19 vs other viruses), classifier D (Covid-19 vs bacteria), or classifier E (Covid-19 vs healthy).

(B) By focusing on specifically identifying Covid-19 pneumonia and distinguishing it from pneumonia caused by other viruses (classifier C) and bacteria (classifier D), our AI models achieved 99.95% diagnostic accuracy for Covid-19 from reading chest X-ray images of virus-caused pneumonia and 99.85% accuracy for Covid-19 from reading the images of pneumonia caused by Covid-19 and bacteria. High diagnostic accuracy was also reached in the groups of classifier A (99.23%), classifier B (99.06%), and Classifier E (100%).

viruses including Covid-19 or bacteria for guiding proper treatments. Therefore, we generated an AI model to distinguish pneumonia from healthy chest X-ray images (classifier A) and another AI model to distinguish virus-caused pneumonia from the one caused by bacteria (classifier B). The model of classifier A achieved 99.23% diagnostic accuracy for pneumonia and the model of classifier B achieved 99.06% accuracy for identifying pneumonia caused by viruses or bacteria (Figure 3B).

Table 1. Individual CNN is less sufficient than 17 CNNs in achieving a clinical-grade diagnostic accuracy for viral and bacterial pneumonia

CNNs	Diagnostic accuracy (%)					Average (A + B + C + D + E)/5
	Classifier A	Classifier B	Classifier C	Classifier D	Classifier E	
AlexNet	94.38	82.87	99.49	99.70	97.78	94.84
GoogleNet	94.48	79.64	98.13	99.70	98.10	94.01
Vgg16	94.10	79.87	99.63	99.70	98.41	94.34
ResNet18	94.00	81.58	99.07	98.80	98.89	94.47
SqueezeNet	94.38	80.09	99.53	100.00	98.25	94.45
MobileNetv2	93.62	79.87	99.07	99.10	96.98	93.73
Inceptionv3	93.14	85.01	99.07	98.49	98.25	94.79
DenseNet201	93.81	80.30	99.53	99.70	97.62	94.19
Xception	95.52	82.01	99.53	99.10	96.35	94.50
Vgg19	94.48	79.66	99.07	100.00	98.89	94.42
Places365GoogleNet	93.71	84.15	99.53	99.40	97.78	94.91
InceptionResNetv2	91.14	82.66	98.60	99.40	97.30	93.82
ResNet50	94.00	79.23	99.53	98.49	97.30	93.71
ResNet101	93.05	77.09	98.13	99.40	97.14	92.96
NASNetMobile	94.76	79.23	99.53	98.80	96.67	93.80
NASNetLarge	93.05	84.58	99.07	99.40	96.17	94.45
ShuffleNet	94.10	80.09	99.07	99.10	97.46	93.96
Our Platform (with combined 17 CNNs)	99.23	99.06	99.95	99.85	100.00	99.62

We originally developed the 17 CNN deep learning platform (Li et al., 2020a) and used it in this study. To demonstrate that our 17 CNN approach is more superior than any individual CNNs often used in the AI field, we analyzed all five classification groups using each of the 17 CNNs separately and achieved a diagnostic accuracy ranging between 79 and 99%, contrasting sharply with the diagnostic accuracy ranging between 99.06 and 100% accuracy using our combined 17 CNN platform (Table 1). It is necessary to point out that when all classifiers (A–E) were analyzed by the same CNN, no individual CNN alone could achieve a diagnostic accuracy greater than 99% (Table 1). By contrast, our combined 17 CNN platform achieved greater than 99% of diagnostic accuracy for all classifiers (Table 1). In our study on AI-assisted pneumonia diagnosis for Covid-19 detection, we built several classifiers, and the classification outcomes with those classifiers shown in the confusion matrices also reflected the high accuracy of our combined multiple CNN deep learning platform (Figure 4).

We should point out that accuracy only measures the number of correctly predicted values among the total predicted value. Although it is a good measure of performance, it is not complete and does not work well when the cost of false negatives is high. To further evaluate our deep learning platform, we employed more evaluation measures including precision (PPV: positive predictive value), NPV (negative predictive value), recall (sensitivity), specificity, and F1 score, because these methods are believed to be valuable ways for validating performance evaluation measures (Powers, 2008; Tharwat, 2021). Our deep learning platform allowed us to obtain high values in precision (>99%), negative predictive value (>98%), recall (>98%), specificity (>99%), and F1 score (>98%) (Table 2).

DISCUSSION

Besides chest X-ray, lung CT images are also taken for clinical diagnosis of Covid-19 pneumonia. Because the availability of CT examination is often limited to larger hospitals, we focused on developing our AI models by solely using chest X-ray images that can be obtained from almost any medical facility, including small clinics/hospitals, even in the remote areas of countries.

In this study, we used transfer learning which allows transferring knowledge from one domain to another by using trained weights from the previous domain. Traditionally, the weight matrices of several layers in a



Figure 4. Confusion matrices for binary classifications A, B, C, D, and E

The green quadrants summarize the correct classifications made by the system, and the red quadrants summarize the incorrect classifications made by the system. TP: true positives; FP: false positives; TN: true negatives; FN: false negatives.

CNN are initially frozen while training on the secondary domain and only the remaining layers are fine-tuned. This process works well when an overlapping region in the low-level features is shared by both domains. In our case, because the ImageNet and the COVID-19 datasets belong to nonoverlapping domains, the trained weights from the ImageNet dataset were used to initialize the weights of our model, and none of them were frozen.

When multiple models are being used to classify a single X-ray image, the final classified class is decided by a majority rule (May, 1952). Majority rule is a decision rule that selects alternatives which have a majority, i.e., maximum votes among those models involved. The idea has been introduced in this study from one of the election theories, called approval voting. Under approval voting, a voter indicates which candidate he or she approves. A candidate receives one point for each voter that approves the candidate. A candidate receives no points for each voter that does not approve the candidate. For a single candidate election, the candidate with the most points wins the election. Naturally, approving of all candidates or disapproving of all candidates does not change the difference in the number of points the candidates receive. If there are an odd number of voters and no voter approves or disapproves of both candidates, then approval voting is equivalent to majority rule: each voter gives one point to the candidate that he or she prefers and the candidate with majority of the points wins the election. Determining a winner for a two-candidate election is easy, which will be a binary classification problem. It has been shown that the majority rule is the only two-candidate election procedure in which each voter is treated equally, that is, only the number of votes matters, not who casts the votes; each candidate is treated equally, that is, only the number of votes that a candidate receives determines if he or she wins the election; besides, a candidate can never be harmed by receiving more votes, that is, if a candidate wins the election, then they would still win the election if some of the voters who had voted for the candidate's opponent now voted for the candidate (May, 1952).

In general, the quality of available chest X-ray images of patients varies hugely across hospitals, and it is challenging to generate highly accurate AI models for diagnosing Covid-19 pneumonia. In some published studies on deep learning, the diagnostic accuracy of about 90% for Covid-19 with a relatively small number of cases implies that a significant number of false negative cases existed in the data sets (Jin et al., 2020; Li et al., 2020b). Practically, we believe that an accuracy close to 100% is required for Covid-19 diagnosis in a clinical setting. Recently, several studies suggest that the use of chest X-ray images may help to assess the severity of Covid-19 (Cohen et al., 2020; Wong et al., 2021; Zhu et al., 2020), emphasizing the clinical significance of chest X-ray in diagnosing Covid-19 pneumonia. In our study, we have developed reliable AI models with nearly 100% diagnostic accuracy for Covid-19 pneumonia by solely reading chest X-ray images of patients, building a solid foundation for using the models in the clinic.

Table 2. Statistical analysis of performance of five binary classifiers

Measures	Classifier A pneumonia vs normal	Classifier B virus vs bacteria	Classifier C Covid-19 vs other virus	Classifier D Covid-19 vs bacteria	Classifier E Covid-19 vs normal
Accuracy	0.9923	0.9906	0.9995	0.9985	1.0
Precision (PPV)	0.9994	0.9916	1.0	0.9975	1.0
Negative predictive value (NPV)	0.9896	0.9898	0.9993	0.9988	1.0
Recall (Sensitivity)	0.9876	0.9879	0.9987	0.9962	1.0
Specificity	0.9995	0.9929	1.0	0.9992	1.0
F1 Score	0.9935	0.9897	0.9993	0.9968	1.0

We have trained and built five binary classification models called Classifier A, Classifier B, Classifier C, Classifier D, and Classifier E (Figures 2 and 3). Our AI models were built mainly for diagnosing Covid-19 but were also capable of identifying non-Covid-19 pneumonia caused by other types of virus or bacteria, providing an opportunity to expand the use of our models to diagnose non-Covid-19 viral or bacterial pneumonia. This approach is meaningful because treatment options for viral and bacterial pneumonia are different. We envision that when individuals who have had a lung infection but otherwise healthy visit to the clinic, our AI models can help to read chest X-ray images to determine whether they have had any form of pneumonia caused by viruses or bacteria, followed by confirming whether they have had Covid-19 pneumonia.

In summary, we took a binary classification deep learning approach using our combined 17 CNNs and core voting algorithm by reading whole chest X-ray images and classifying them as either positive or negative to Covid-19. As a result, we have achieved nearly 100% diagnostic accuracy for Covid-19 pneumonia with high sensitivity and specificity. Our immediate next step would be to apply our AI models in a clinical trial for chest X-ray-based diagnosis of Covid-19.

Limitations of the study

As we showed in our results, we have achieved a high accuracy in identifying Covid-19 pneumonia using our deep learning method for potential clinical use. However, we could not explain why the accuracy did not reach 100% in some comparison groups. In other words, we do not know whether we need to further improve our deep learning method or to verify the correctness of image labeling in the public datasets, although the latter is obviously impossible to achieve. Before clinical use of our deep learning method for diagnosing Covid-19 pneumonia, we may need to ensure a control of the image collection process to avoid possible mislabeling of any chest X-ray images.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Material availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
 - Image datasets
 - Core voting algorithm
- [METHOD DETAILS](#)
 - Hardware and software
 - Deep learning networks and convolutional neural networks (DCNNs)
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (R01CA176179, R01CA222590) to S.L.

AUTHOR CONTRIBUTIONS

D.L. and S. L. conceived the study, obtained chest X-ray images, and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 13, 2021

Revised: January 20, 2022

Accepted: March 2, 2022

Published: April 15, 2022

REFERENCES

- Appenzeller, T. (2017). The scientists' apprentice. *Science* 357, 16–17.
- Campanella, G., Hanna, M.G., Geneslaw, L., Miralflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., and Fuchs, T.J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* 25, 1301–1309.
- Cohen, J.P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A.F., Shen, B., Mahsa, H.K., Ghassemi, M., Li, H., et al. (2020). Predicting COVID-19 pneumonia severity on chest X-ray with deep learning. *Cureus* 12, e9448.
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368, 779–782.
- Grubaugh, N.D., Hanage, W.P., and Rasmussen, A.L. (2020). Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell* 182, 794–795. <http://image-net.org/about-overview>, (2016). ImageNet Overview.
- Hu, J., Li, C., Wang, S., Li, T., and Zhang, H. (2020). Genetic Variants Are Identified to Increase Risk of COVID-19 Related Mortality from UK Biobank Data. Preprint at medRxiv. <https://doi.org/10.1101/2020.11.05.20226761>.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 395, 497–506.
- Hussain, L., Nguyen, T., Li, H., Abbasi, A.A., Lone, K.J., Zhao, Z., Zaib, M., Chen, A., and Duong, T.Q. (2020). Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection. *Biomed. Eng. Online* 19, 88.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y.X., Lu, P.X., and Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med. Surg.* 4, 475–477.
- Jin, C., Chen, W., Cao, Y., Xu, Z., Tan, Z., Zhang, X., Deng, L., Zheng, C., Zhou, J., Shi, H., et al. (2020). Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat. Commun.* 11, 5088.
- Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C.S., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172, 1122–1131.e29.
- Kikkiseti, S., Zhu, J., Shen, B., Li, H., and Duong, T.Q. (2020). Deep-learning convolutional neural networks with transfer learning accurately classify COVID-19 lung infection on portable chest radiographs. *Peer J.* 8, e10309.
- Kirby, T. (2021). New variant of SARS-CoV-2 in UK causes surge of COVID-19. *Lancet Respir. Med.* 9, e20–e21.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827.e19.
- Lam, T.T., Jia, N., Zhang, Y.W., Shum, M.H., Jiang, J.F., Zhu, H.C., Tong, Y.G., Shi, Y.X., Ni, X.B., Liao, Y.S., et al. (2020). Identifying SARS-CoV-2-related coronaviruses in malayan pangolins. *Nature* 583, 282–285.
- LeCun, Y., and Bengio, Y. (1998). Convolutional networks for images, speech, and time-series. In *The Handbook of Brain Theory and Neural Networks*, M.A. Arbib, ed. (MIT Press), pp. 255–258.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.
- Li, D., Bledsoe, J.R., Zeng, Y., Liu, W., Hu, Y., Bi, K., Liang, A., and Li, S. (2020a). A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. *Nat. Commun.* 11, 6004.
- Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., et al. (2020b). Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. *Radiology* 296, E65–E71.
- May, K.O. (1952). A set of independent necessary and sufficient conditions for simple majority decisions. *Econometrica* 20, 680–684.
- Medicine, U.S.N.L.o (2020). Tuberculosis chest X-Ray image data sets. <https://ceb.nlm.nih.gov/tuberculosis-chest-X-ray-image-datasets/>.
- Paul, C.J. (2020). Covid-19 image data collection. <https://github.com/ieee8023/covid-chestxray-dataset/>.
- Powers, D.M.W. (2008). Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation. *J. Machine Learn. Tech.* 2, 37–63.
- Razavian, N. (2019). Augmented reality microscopes for cancer histopathology. *Nat. Med.* 25, 1334–1336.
- Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117.
- Shen, B., Hoshmand-Kochi, M., Abbasi, A., Glass, S., Jiang, Z., Singer, A.J., Thode, H.C., Li, H., Hou, W., and Duong, T.Q. (2021). Initial chest radiograph scores inform COVID-19 status, intensive care unit admission and need for mechanical ventilation. *Clin. Radiol.* 76, 473.e71–473.e77.
- Tang, J.W., Tambyah, P.A., and Hui, D.S. (2020). Emergence of a new SARS-CoV-2 variant in the UK. *J. Infect.* 82, e27–e28.
- Tharwat, A. (2021). Classification assessment methods. *Appl. Comput. Inform.* 17, 168–192.
- Wang, T., Du, Z., Zhu, F., Cao, Z., An, Y., Gao, Y., and Jiang, B. (2020). Comorbidities and multi-organ injuries in the treatment of COVID-19. *Lancet* 395, e52.
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R.M. (2017). ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and location of common thorax diseases.

Preprint at arXiv. <https://doi.org/10.48550/arXiv.1705.02315>.

Weiss, K., Khoshgoftaar, T., and Wang, D. (2016). A survey of transfer learning. *J Big Data* 3, 9.

Wong, A., Lin, Z.Q., Wang, L., Chung, A.G., Shen, B., Abbasi, A., Hoshmand-Kochi, M., and Duong, T.Q. (2021). Towards computer-aided severity assessment via deep neural networks for geographic and opacity extent scoring of SARS-CoV-2 chest X-rays. *Sci. Rep.* 11, 9315.

Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L.,

et al. (2020). Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* 8, 420–422.

Zhang, H., Zhou, P., Wei, Y., Yue, H., Wang, Y., Hu, M., Zhang, S., Cao, T., Yang, C., Li, M., et al. (2020a). Histopathologic changes and SARS-CoV-2 immunostaining in the lung of a patient with COVID-19. *Ann. Intern. Med.* 172, 629–632.

Zhang, J.J., Cao, Y.Y., Dong, X., Wang, B.C., Liao, M.Y., Lin, J., Yan, Y.Q., Akdis, C.A., and Gao, Y.D. (2020b). Distinct characteristics of COVID-19 patients with initial rRT-PCR-positive and rRT-

PCR-negative results for SARS-CoV-2. *Allergy* 75, 1809–1812.

Zhu, J., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Li, H., and Duong, T.Q. (2020). Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PLoS One* 15, e0236621.

Ziegler, C.G.K., Allon, S.J., Nyquist, S.K., Mbanjo, I.M., Miao, V.N., Tzouanas, C.N., Cao, Y., Yousif, A.S., Bals, J., Hauser, B.M., et al. (2020). SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* 181, 1016–1035.e19.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB	Mathworks	https://www.mathworks.com/
Core Voting Algorithms with multiple CNNs	Nature Communications	https://www.nature.com/articles/s41467-020-19,817-3
Code (Classifier A, B, C, D and E)	This study	https://fts.umassmed.edu/ (user name: dli; password: Dong2022)
Pretrained CNN Models	Mathworks	https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html
Other		
Model Evaluation Measures	Wikipedia	https://en.wikipedia.org/wiki/F-score
X-ray Image Datasets (Healthy, Bacteria, Virus and Covid-19)	This study	https://fts.umassmed.edu/ (user name: dli; password: Dong2022)

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Shaoguang Li (shaoguang.li@umassmed.edu).

Material availability

We have prepared four image datasets (Normal or Healthy, Bacteria, Other virus, and Covid-19) in this study. The datasets used to train and evaluate the proposed platform is comprised of a total 10,182 chest X-ray images, and these images are available from <https://fts.umassmed.edu> (user name: dli; password: Dong2022).

Data and code availability

- Whole source code can be found from <https://fts.umassmed.edu> (user name: dli; password: Dong2022) or obtained by sending a request to shaoguang.li@umassmed.edu.
- This paper was produced using large volumes of publicly available image data. The authors have made every effort to make available links to these resources as well as the software methods and information used to produce the dataset, analyses, and summary information. The datasets used in this study are available online (<https://fts.umassmed.edu/>).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Image datasets

All Covid-19 chest X-ray images were obtained from a publicly-available depository site (<https://github.com/ieee8023/covid-chestxray-dataset/tree/master/images>). They are real cases for patients who tested positive for Covid-19 in hospitals across the global. Non-Covid-19 chest X-ray images were obtained from the Kaggle's Chest X-ray Images (Pneumonia) dataset (<https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>). Based on the causes of pneumonia, we grouped all cases as either Covid-19 or Non-Covid19 (healthy, bacterial pneumonia and other viral pneumonia).

The dataset used to train and evaluate the proposed platform is comprised of a total 10,182 chest X-ray images, and these images are available from <https://github.com/ieee8023/covid-chestxray-dataset/tree/master/images> and <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>. We combined and modified several different datasets that are publicly available (Kermany et al., 2018; Medicine, 2020; Paul, 2020; Wang et al., 2017). There is also more information related to those images (Cohen et al., 2020; Jaeger et al., 2014).

Core voting algorithm

Like many different research groups, we have used different CNNs individually but diagnostic accuracy has not been satisfactory. In our study, we initially used each of the 17 CNNs, respectively, to analyze the chest X-ray images of Covid-19 in each of the five classification groups, and found that the average diagnostic accuracy by using one CNN was ranged from 79% to 99% (Table 1). In our view, the diagnostic accuracy needs to be near 100% or greater than 99% prior to employing any deep learning model in medical practice. This is why we programmed multiple models (17 CNNs) into one system with our core algorithms to enhance the performance of deep learning with a goal of achieving 100% diagnostic accuracy. As a result, we have indeed reached 100% accuracy, which is superb to a sole use of any one of the 17 CNNs. In fact, we were not particularly interested in which network(s) contribute the most/the best to the output, because our model was treated as a black box, which is how deep learning should work. However, what we do know is that the output of combining the 17 CNNs is much better than any individual network.

In order to get the multiple CNNs to work together, a core algorithm is developed based on a voting mechanism. In the classification process, each individual CNN model votes for the Covid-19 results either Yes or No. Yes will get a score +1 and No will get a score −1. An array Vote(n) is created after receiving all of 17 CNNs contributing vote scores. Then the final classification output (Covid or Non-Covid) can be calculated by adding up all of the scores of 17 CNNs.

$$\text{Classification} = \sum_{n=1}^{17} \text{Vote}(n) \quad (\text{Equation 1})$$

Covid-19, if Classification ≥ 0

Non-Covid-19, if Classification < 0

Our ability to combine 17 CNNs and use them together as a single model is definitely unprecedented. This single model has all of the layers built in those 17 CNNs for conducting transfer deep learning with our datasets, and this novel approach allowed us to achieve a high diagnostic accuracy for Covid-19.

METHOD DETAILS

Hardware and software

CPU Sever and computer used for conducting all experiments are described previously (Li et al., 2020a). Briefly, the MATLAB2019a was used for training AI models. In data preparation, programming and deployment, the toolboxes provided by MATLAB were used, including the deep-learning toolbox and the image processing toolbox.

Deep learning networks and convolutional neural networks (DCNNs)

In this study, we used deep learning to generate computational models that are composed of multiple processing layers, including convolution, activation or ReLU, and pooling (Figure 2B). Because the size of the publicly-available COVID-19 dataset is relatively small compared to standard datasets used in deep learning, we applied transfer learning to augment the decision-making process, and the pre-trained networks on ImageNet are deep CNNs originally designed to classify images in 1,000 categories (<http://image-net.org/about-overview>, 2016). We reused the network architecture of the CNN to classify those X-ray images in two categories, such as Healthy vs Covid-19, Virus vs Covid-19, Bacteria vs Covid-19, and Virus vs Bacteria, from several data sources. Then, we determined the type of classification technique that could be applied for distinguishing the two classes. Based on the collected images, we could identify pre-processing techniques that would assist our classification process. We could also determine the type of CNN architecture utilized for this study based on the similarities within the class and differences across classes.

Each every of 17 CNNs was trained for many iterations (approximately 20–30 epochs with batch sizes ranging from 16 to 64, learning rate from 0.0001 to 0.001, validation frequency is 20) before convergence. The detail procedure of training is illustrated in Figure 3B. We split our data into 3 datasets. 10% of the data were used for validation, 10% of the data were reserved for testing, while the remaining 80% of the data were used for training. During training, validation data is useful to detect if the network is overfitting. All of 17 trained CNNs were incorporated into a core voting algorithm to work out the final classification output (Figure 2A) with comparison to the performance of each of individual CNNs (Table 1).

Diagnostic accuracy was used as a measure to evaluate the diagnostic performance, which involved in the use of the following terms: true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Diagnostic accuracy was calculated as the following:

$$\text{Diagnostic accuracy} = (TP + TN) / (TP + FN + FP + TN) \quad (\text{Equation 2})$$

To further evaluate our deep learning platform, we employed more evaluation measures including the following:

$$\text{Precision (positive predictive value, PPV)} = TP / (TP + FP) \quad (\text{Equation 3})$$

$$\text{NPV (negative predictive value)} = TN / (TN + FN) \quad (\text{Equation 4})$$

$$\text{Recall (sensitivity)} = TP / (TP + FN) \quad (\text{Equation 5})$$

$$\text{Specificity} = TN / (TN + FP) \quad (\text{Equation 6})$$

$$\text{F1 Score} = 2 * [(precision * recall) / (precision + recall)] \quad (\text{Equation 7})$$

QUANTIFICATION AND STATISTICAL ANALYSIS

We have calculated a sets of diagnostic testing measures, which is related to the field of binary classification (<https://en.wikipedia.org/wiki/F-score>). No other statistical analysis was performed in the study.