RESEARCH

Optimization of vision transformer-based detection of lung diseases from chest X-ray images

Jinsol Ko[1,2†], Soyeon Park[3†] and Hyun Goo Woo[1,2*]

Abstract

Background Recent advances in Vision Transformer (ViT)-based deep learning have significantly improved the accuracy of lung disease prediction from chest X-ray images. However, limited research exists on comparing the effectiveness of different optimizers for lung disease prediction within ViT models. This study aims to systematically evaluate and compare the performance of various optimization methods for ViT-based models in predicting lung diseases from chest X-ray images.

Methods This study utilized a chest X-ray image dataset comprising 19,003 images containing both normal cases and six lung diseases: COVID-19, Viral Pneumonia, Bacterial Pneumonia, Middle East Respiratory Syndrome (MERS), Severe Acute Respiratory Syndrome (SARS), and Tuberculosis. Each ViT model (ViT, FastViT, and CrossViT) was individually trained with each optimization method (Adam, AdamW, NAdam, RAdam, SGDW, and Momentum) to assess their performance in lung disease prediction.

Results When tested with ViT on the dataset with balanced-sample sized classes, RAdam demonstrated superior accuracy compared to other optimizers, achieving 95.87%. In the dataset with imbalanced sample size, FastViT with NAdam achieved the best performance with an accuracy of 97.63%.

Conclusions We provide comprehensive optimization strategies for developing ViT-based model architectures, which can enhance the performance of these models for lung disease prediction from chest X-ray images.

Keywords Optimizer, Vision transformer, Lung disease, X-ray

Background

Deep learning algorithms have demonstrated remarkable success in distinguishing lung diseases by analyzing chest X-rays. Convolutional neural networks (CNNs), in particular, have achieved promising results in this domain [1–4]. However, the introduction of Vision Transformers (ViTs) has opened up new avenues for lung disease classification [5]. Unlike CNNs, ViTs utilize self-attention mechanisms, converting images into sequences of image patches that are then processed by a transformer [6]. This novel approach has yielded state-of-the-art performance in various computer vision tasks, including image classification, object detection, and segmentation. Indeed, ViTs have surpassed CNNs in detecting COVID-19 from chest X-rays, achieving accuracies above 96% [7–10]. Moreover, new ViT-based models, such as FastViT and CrossViT, have also emerged and exhibited promising results. FastViT is a cutting-edge hybrid form of ViT that strikes an optimal

†Jinsol Ko and Soyeon Park contributed equally to this work.

*Correspondence:
Hyun Goo Woo
hg@ajou.ac.kr
[1] Department of Physiology, Ajou University School of Medicine, Suwon, Republic of Korea
[2] Department of Biomedical Science, Graduate School, Ajou University, Suwon, Republic of Korea
[3] Ajou University School of Medicine, Suwon, Republic of Korea

Ko *et al. BMC Medical Informatics and Decision Making*   (2024) 24:191

Page 2 of 8

balance between latency and accuracy [11]. CrossViT employs a dual-branch transformer with a cross-attention mechanism, generating stronger image features [12].

On the other hand, enhancing the accuracy of deep learning models remains a crucial area of research. Data augmentation has proven to be an effective technique for improving neural network precision. However, not all models benefit equally from augmentation, necessitating the exploration of broadly applicable methods such as loss function optimization [13]. Indeed, various optimization techniques have demonstrated varying performances in training CNN models. For instance, the Xception model, which was pretrained for classifying chest X-rays into normal, COVID-19, and pneumonia categories, achieved the highest accuracy with the Root Mean Square Propagation (RMSProp) optimizer [14]. Meanwhile, other studies have reported superior performance with the Adaptive Moment Estimation (Adam) optimizer [15, 16] or Adaptive Gradient (AdaGrad) optimizer [17] when coupled with CNN models. These findings underscore the model-dependent nature of optimizer performance. Similarly, ViT models exhibit varying performance depending on the chosen optimizer. Rectified Adam (RAdam) [8, 9], Adam with cosine decay [7], NovoGrad [10], and AdaBelief [18] have been applied to ViTs. Notwithstanding, a comprehensive comparison of optimization methods for ViT models in chest X-ray image classification has not been conducted.

This study aimed to identify the best performing optimization method in training ViT-based models for predicting lung diseases from chest X-ray images. We evaluated the following six optimization methods known to produce promising results in computer vision: Adam, Adam with weight decay (AdamW), Nesterov accelerated Adam (NAdam), RAdam, Stochastic Gradient Descent with weight decay (SGDW), and Momentum. Stochastic Gradient Descent (SGD), a cornerstone optimizer in modern neural networks, paved the way for achieving minimal loss values [19, 20]. Enhancements to SGD, such as Momentum and Nesterov acceleration techniques, have improved its effectiveness, resulting in faster convergence and higher accuracy with fewer steps [13]. In addition, several advanced algorithms, such as AdaGrad, RMSProp, Adadelta, and Adam, have been proposed to effectively navigate complex functions with local extremes. The Adam-based algorithms utilize exponential moving averages and excel at refining minimization processes in functions with multiple extremes. A notable advantage of these methodologies lies in their ability to derive corrected estimates for bias, effectively countering the effects of initial bias settings [13]. Moreover, building upon the success of the Adam optimizer, several variants, including AdamW, NAdam, and RAdam, have emerged

as promising alternatives. These variants introduce subtle refinements to the original algorithm, incorporating features such as weight decay or adaptive learning rate adjustments to further enhance its performance and robustness. In this study, we conducted a comprehensive comparison of these optimization algorithms, including Adam and its variants (AdamW, NAdam, and RAdam) and the traditional Momentum and SGDW (a variant of SGD). This in-depth evaluation could address the best fitted optimizer according to the model and dataset.

Methods

Dataset

We obtained a publicly available chest X-ray dataset from an open-source repository, details of which can be found in the availability of data and materials section [21–31]. The dataset comprised 19,003 images, categorized into six disease classes: COVID-19 ($n = 3{,}616$), Viral Pneumonia ($n = 1{,}345$), Bacterial Pneumonia ($n = 2{,}772$), Middle East Respiratory Syndrome (MERS, $n = 144$), Severe Acute Respiratory Syndrome (SARS, $n = 134$), and Tuberculosis ($n = 800$). Additionally, the dataset included normal chest X-rays ($n = 10{,}192$) for reference.

Model structure

The RGB values of the image data were normalized to an average of 0.5 and a standard deviation of 0.5. The image size was set to 224 x 224 pixels, and each image was divided into 16 x16 patches, which were then flattened and linearly projected to create patch embeddings. These embeddings were combined with position embeddings to preserve positional information. We utilized the ViT-B/16 model [5], which was pretrained on ImageNet data [32]. For each block, layer normalization and residual connection were applied [5]. Following recommendations from previous research [33], we trained the model with a batch size of 32 for 15 epochs, shuffling the dataset before each epoch.

We applied six different optimizers, including Adam, AdamW, NAdam, RAdam, SGDW, and Momentum (Supplementary Table 1). Each optimizer was tested on three different learning rates: 10^{-4}, 10^{-5}, and 10^{-6}. All the parameters used in the models are summarized in Supplementary Table 2.

Evaluation

The performance of each model was evaluated using four key metrics: accuracy, F1-score, precision, and recall [34]. Accuracy, precision, and recall are parameters that can effectively assess the performance of a model, and F1-score is known to be robust against data with imbalanced sample size [34, 35].

$$\text{Accuracy} = (\text{True positive} + \text{True negative})/(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})$$

$$\text{F1-score} = \text{Harmonic mean of precision and recall} = 2 \text{ x Precision x Recall} /(\text{Precision} + \text{Recall})$$

$$\text{Precision} = \text{Positive Predictive Value} = \text{True positive} /(\text{True positive} + \text{False positive})$$

$$\text{Recall} = \text{Sensitivity} = \text{True positive} /(\text{True positive} + \text{False negative})$$

Results

Classification of the overall classes with various models and optimizers

The sample sizes for each class in the chest X-ray dataset ranged from 134 to 10,192 images. The presence of imbalanced sample sizes may compromise the learning process, leading to biased outcomes [36, 37]. Thus, we evaluated the effect of the imbalanced sample size by performing analysis using the 4 class dataset with sample sizes greater than 1,000 (i.e., Normal, COVID-19, Viral Pneumonia, and Bacterial Pneumonia). We compared these results with those from the complete dataset also including the small sample-sized classes (i.e., MERS, SARS, and Tuberculosis). The dataset was randomly split into the training dataset (80%) and test dataset (20%), and the model performance was evaluated with different learning rates of 10^{-4}, 10^{-5}, and 10^{-6}. The overall structure of the model is demonstrated in Fig. 1.

First, we evaluated the performance of the models with no optimizer, which revealed poor accuracies in each test of the ViT with 4 class (30.90%) and the ViT with 7 class (14.89%), respectively (Supplementary Fig. 1). FastViT (6.10%) and CrossViT (22.72%) with 7 class also showed poor performance, indicating the need for the use of an optimizer (Supplementary Fig. 2). Next, we evaluated the performance of the ViT model with different optimizers and learning rates (i.e., 10^{-4}, 10^{-5}, and 10^{-6}) using the 4 class dataset. In general, ViT exhibited robust classification independent of the utilization of different optimizers (Fig. 2a, Supplementary Table 3, and Supplementary Fig. 3). However, SGDW and Momentum exhibited relatively lower accuracies (SGDW: 81.71%, Momentum: 82.18%) at learning rate 10^{-6} compared to the other optimizers (>86%). Remarkably, RAdam showed the highest accuracy (95.87%, learning rate 10^{-5}), while Adam achieved the highest F1-score (94.71%, learning rate 10^{-5}). Adam-based optimizers (Adam, AdamW, NAdam, and RAdam) consistently demonstrated superior performance across all metrics, outperforming other optimizers (SGDW and Momentum). This finding

could be attributed to the adaptive momentum algorithm employed by Adam-based optimizers. In addition, we evaluated the performance of ViT with the imbalanced sample-sized dataset (i.e., 7 class dataset). RAdam at learning rate 10^{-5} demonstrated the highest accuracy (96.61%) and F1-score (96.62%) in the 7 class dataset (Fig. 2b, Supplementary Table 3, and Supplementary Fig. 4).

In addition to ViT, when we evaluated the performance of FastViT and CrossViT, FastViT showed best performance with NAdam (accuracy 97.63%, F1-score 97.64%, learning rate 10^{-4}; Fig. 2c, Supplementary Table 3, and Supplementary Fig. 5), while CrossViT showed best performance with AdamW (accuracy 96.95%, F1-score 96.94%, learning rate 10^{-5}; Fig. 2d, Supplementary Table 3, and Supplementary Fig. 6). These results indicate that Adam-based optimizers perform well in ViT-based models. Comparing results of the imbalanced 7 class dataset in all three models (ViT, FastViT, and CrossViT), both the highest accuracy and F1-score were achieved by FastViT with NAdam. This finding may indicate the robustness of NAdam against sample imbalance.

The disease class prediction with various ViT models and optimizers

Next, we sought to evaluate whether the prediction performance for each disease class varied depending on the models and optimizers by calculating F1-scores. In the 4 class classification by ViT, the Normal class exhibited robust prediction performance across all optimizers, showing the highest F1-score with RAdam (98.05%, learning rate 10^{-5}), whereas the Viral Pneumonia class with NAdam showed the lowest performance (52.45%, learning rate 10^{-4}) (Fig. 3a and Supplementary Table 4). In addition, we also evaluated the performance of ViT with varying optimizers in the 7 class classification. Similarly, RAdam showed satisfactory performance for predicting the Normal class (98.73%, learning rate 10^{-5}), while AdamW was the best optimizer for the class of Tuberculosis (99.07%, learning rate 10^{-6}) (Fig. 3b and Supplementary Table 4). In FastViT and CrossViT, Tuberculosis was
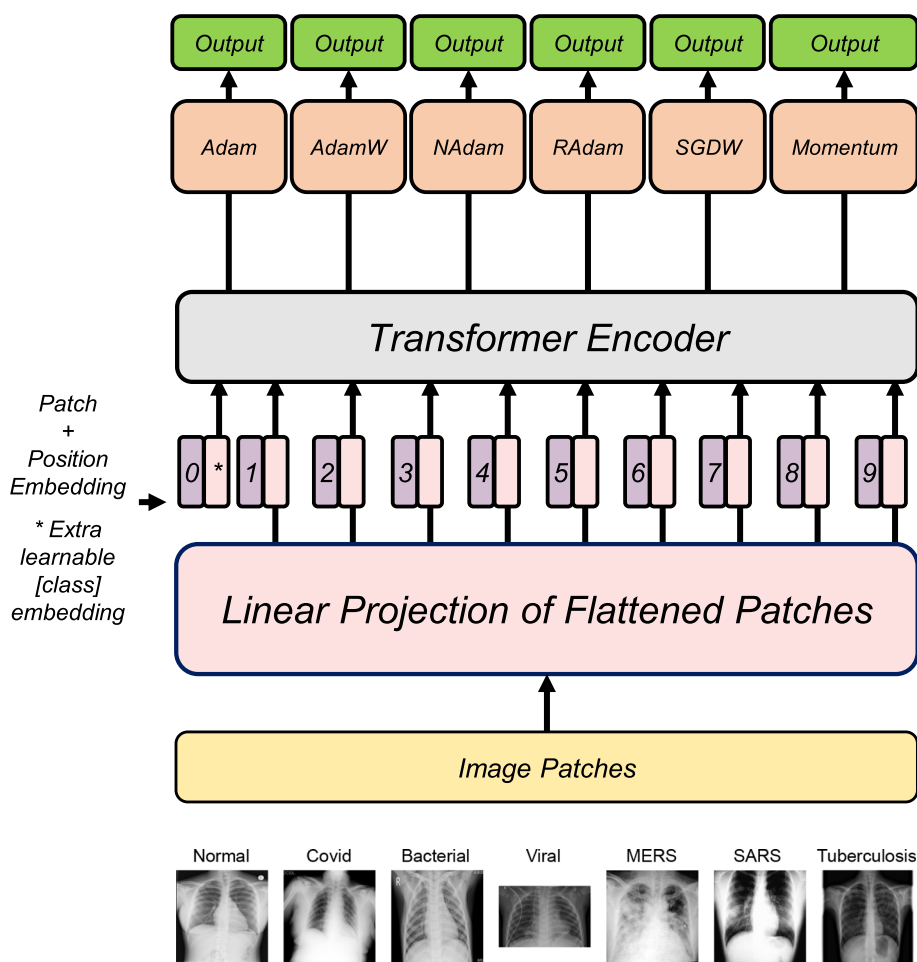
Fig. 1 Schematic overview of the analysis workflow

the best predicted class (FastViT: Adam, 100%, learning rate 10^{-4}; CrossViT: NAdam, 99.69%, learning rate 10^{-5}) in both models (Fig. 3c, d and Supplementary Table 4).

Discussion

In this study, we evaluated the performance of different optimizers in ViT-based prediction of lung diseases from chest X-ray images. By comparing the performance of six different optimizers, we found that Adam-based optimizers displayed superior performance compared to SGDW and Momentum. This might be because Adam-based optimizers are implemented with an adaptive momentum, adapting learning rates to the parameters. Similarly, a previous study has demonstrated that SGD and Momentum increased error rates with the learning rate lower than 10^{-5} in CNN-based classification of brain MRI [38]. These results consistently suggest that Adam-based optimizers are well suited for ViT-based models in predicting lung diseases using chest X-rays.

We demonstrated that RAdam showed optimal performance in predicting the balanced 4 class dataset in the ViT model. Supporting this, RAdam has shown satisfactory performance in ViT-based COVID-19 classification, although Adam or RMSProp did not [8, 9]. RAdam is a recent technique that combines the strengths of both Adam and SGD, ensuring swift convergence without easily succumbing to local optima. RAdam rectifies the variance of the adaptive learning rate term, which can make the variance consistent. Therefore, the convergence of RAdam is largely unaffected by the initial learning rate value [39].

In the imbalanced 7 class dataset, NAdam in FastViT had the highest accuracy (97.63%), which is comparable to previous ViT models for COVID-19 detection [8, 40–42]. NAdam modifies the momentum technique applied in Adam to the Nesterov accelerated gradient (NAG). By combining the advantages of Adam and NAG, NAdam can find the global minimum faster and more accurately than Adam. Unlike Adam, NAdam is based on NAG,
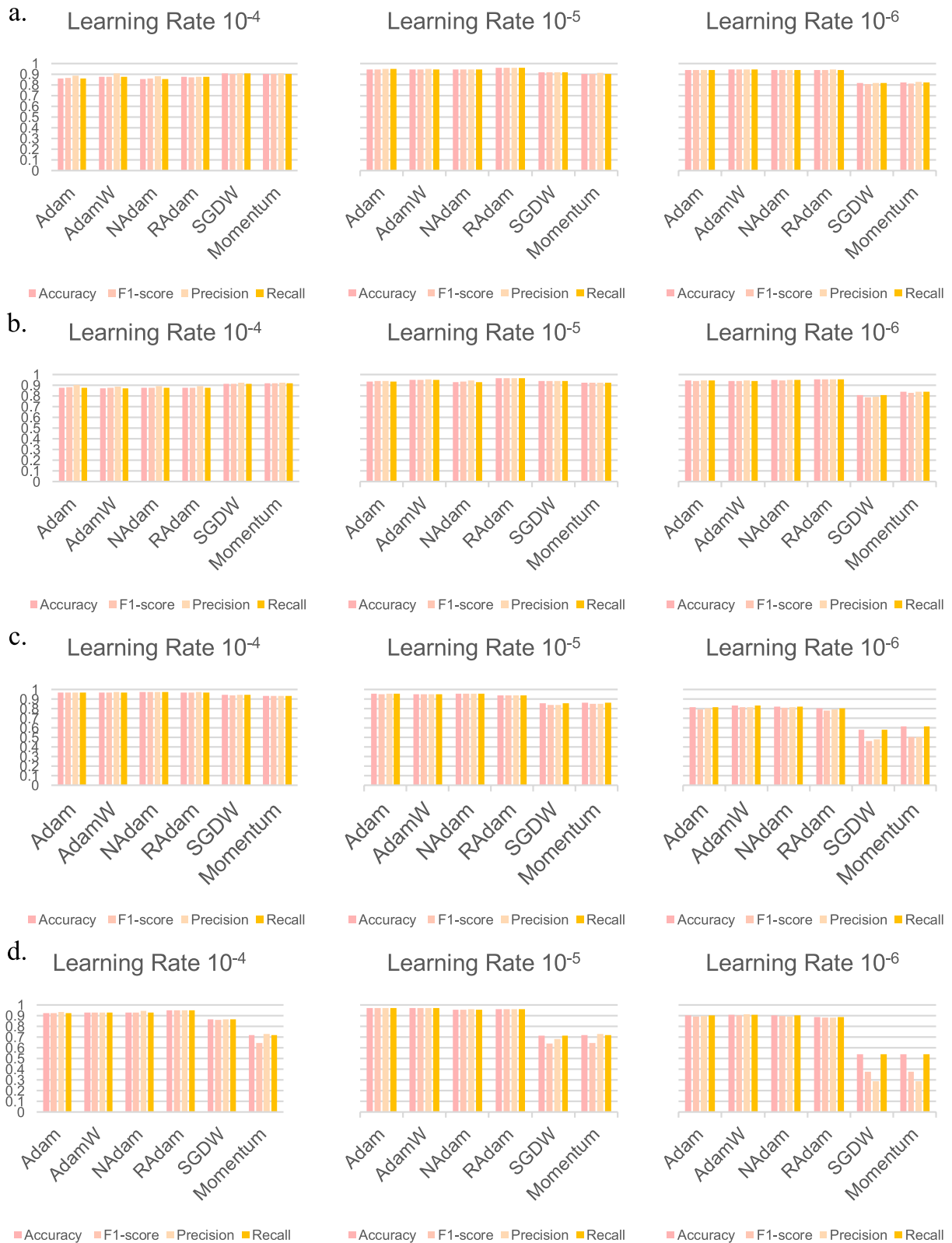
Ko *et al. BMC Medical Informatics and Decision Making* (2024) 24:191

Page 5 of 8

Fig. 2 Classification of the overall classes with various models and optimizers. **A**, **B** The 4 class (**A**) and 7 class (**B**) datasets were classified using the ViT model with various optimizers (Adam, AdamW, NAdam, RAdam, SGDW, and Momentum), respectively. **C**, **D** The 7 class dataset was classified using the FastViT (**C**) or CrossViT (**D**) models with various optimizers, respectively. The evaluation metrics included accuracy, F1-score, precision, and recall, calculated at various learning rates of 10^{-4}, 10^{-5}, and 10^{-6}

Fig. 3 Classification of each disease class with various models and optimizers. **A**, **B** Each class in the 4 class (**A**) and 7 class (**B**) datasets was classified using the ViT model with various optimizers (Adam, AdamW, NAdam, RAdam, SGDW, and Momentum), respectively. **C**, **D** Each class in the 7 class dataset was classified using the FastViT (**C**) or CrossViT (**D**) models with various optimizers, respectively. The evaluation metrics included accuracy, F1-score, precision, and recall, calculated at various learning rates of 10^{-4}, 10^{-5}, and 10^{-6}

Ko *et al. BMC Medical Informatics and Decision Making*   (2024) 24:191

Page 7 of 8

which does not calculate the gradient value at the current position but after moving in the direction of the momentum. Thus, it computes the gradient at a new position [43].

For each class prediction, Normal and Tuberculosis were classified better than other classes. This might be due to the ambiguous feature of X-ray images among COVID-19, Viral Pneumonia, Bacterial Pneumonia, MERS, and SARS classes. The limitation of this study is that the models struggled with detecting small sized classes such as MERS and SARS. Further studies with diverse cases and varied sample sizes and diseases are needed for better performance. Additionally, we only focused on transformer models such as ViT, FastViT, and CrossViT. Future studies may deepen this field of study by investigating other optimizers in various computer vision models, such as hybrid models.

Conclusions

In summary, our analyses of ViT-based models with various optimizers showed varying performance in predicting lung diseases from the chest X-ray images. Adam-based optimizers showed better performance in predicting disease classes. In the balanced dataset, RAdam was the best performing optimizer, while NAdam with FastViT showed the best accuracy in the imbalanced dataset. For the prediction of each disease class, Normal and Tuberculosis were well predicted compared to other classes. Our results might help develop the optimized algorithms with different model architectures and optimizers.

Abbreviations

CNN	Convolutional neural network
ViT	Vision transformer
COVID-19	Coronavirus disease 2019
RMSProp	Root mean square propagation
Adam	Adaptive moment estimation
AdaGrad	Adaptive gradient
RAdam	Rectified Adam
AdamW	Adam with weight decay
NAdam	Nesterov accelerated Adam
SGDW	Stochastic gradient descent with weight decay
SGD	Stochastic gradient descent
MERS	Middle east respiratory syndrome
SARS	Severe acute respiratory syndrome

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02591-3.

Supplementary Material 1.

Supplementary Material 2.

Availability of data and materials

The datasets generated and analyzed during the current study are available in the BIMCV, GitHub, SIRM, eurorad, figshare, and Kaggle repository,
[1] https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711
[2] https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png
[3] https://sirm.org/category/senza-categoria/covid-19/
[4] https://eurorad.org
[5] https://github.com/ieee8023/covid-chestxray-dataset
[6] https://figshare.com/articles/COVID-19_Chest_X-Ray_Image_Repository/12580328
[7] https://github.com/armiro/COVID-CXNet
[8] https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data
[9] https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
[10] https://www.kaggle.com/datasets/057e1b6dc41d9691e59dded4445fa8cc2f0b4b5cbcb49aef9583d95233799d5a
[11] https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed. 2020;196: 105581.
2. Wang L, Lin ZQ, Wong A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep. 2020;10(1):19549.
3. Rahaman MM, Li C, Yao Y, Kulwa F, Rahman MA, Wang Q, Qi S, Kong F, Zhu X, Zhao X. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. J Xray Sci Technol. 2020;28(5):821–39.

Ko *et al. BMC Medical Informatics and Decision Making*   (2024) 24:191

Page 8 of 8

4. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, Van Ginneken B, Madabhushi A, Prince JL, Rueckert D, Summers RM. A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. Proc IEEE. 2021;109(5):820–38.

5. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. arXiv preprint arXiv:201011929.

6. Vaswani A, et al. Attention is all you need. Advances in neural information processing systems. 2017;30. https://doi.org/10.48550/arXiv.1706.03762.

7. Mondal AK, Bhattacharjee A, Singla P, Prathosh AP. xViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography. IEEE J Transl Eng Health Med. 2022;10:1100110.

8. Chetoui M, Akhloufi MA. Explainable Vision Transformers and Radiomics for COVID-19 Detection in Chest X-rays. J Clin Med. 2022;11(11):3013.

9. Krishnan KS, Krishnan KS. Vision transformer based COVID-19 detection using chest X-rays. In: *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC): 2021*: IEEE. 2021. p. 644–8.

10. Shome D, Kar T, Mohanty SN, Tiwari P, Muhammad K, AlTameem A, Zhang Y, Saudagar AKJ. COVID-Transformer: Interpretable COVID-19 Detection Using Vision Transformer for Healthcare. Int J Environ Res Public Health. 2021;18(21):11086.

11. Vasu PKA, Gabriel J, Zhu J, Tuzel O, Ranjan A. FastViT: A Fast Hybrid Vision Transformer using Structural Reparameterization. 2023. arXiv preprint arXiv:230314189.

12. Chen C-FR, Fan Q, Panda R. Crossvit: Cross-attention multi-scale vision transformer for image classification. In: Proceedings of the IEEE/CVF international conference on computer vision: 2021. 2021. p. 357–66.

13. Abdulkadirov R, Lyakhov P, Nagornov N. Survey of Optimization Algorithms in Modern Neural Networks. Mathematics. 2023;11(11):2466.

14. AbdElhamid AA, AbdElhalim E, Mohamed MA, Khalifa F. Multi-Classification of COVID-19 Diagnosis Using Deep Learning Algorithms. Appl Sci. 2022;12(4):2080.

15. Appasami G, Nickolas S. A deep learning-based COVID-19 classification from chest X-ray image: case study. Eur Phys J Spec Top. 2022;231(18):3767–77.

16. Fan Z, Jamil M, Sadiq MT, Huang X, Yu X. Exploiting Multiple Optimizers with Transfer Learning Techniques for the Identification of COVID-19 Patients. J Healthc Eng. 2020;2020(1):8889412. https://doi.org/10.1155/2020/8889412.

17. Shamrat FJM, Azam S, Karim A, Ahmed K, Bui FM, De Boer F. High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images. Comput Biol Med. 2023;155: 106646.

18. Al Rahhal MM, Bazi Y, Jomaa RM, AlShibli A, Alajlan N, Mekhalfi ML, Melgani F. COVID-19 Detection in CT/X-ray Imagery Using Vision Transformers. J Pers Med. 2022;12(2):310.

19. Gastaldi X. Shake-shake regularization. 2017. arXiv preprint arXiv:170507485.

20. Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. Autoaugment: Learning augmentation policies from data. 2018. arXiv preprint arXiv:180509501.

21. RSNA Pneumonia Detection Challenge [https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data].

22. COVID-19 Chest X-Ray Image Repository [https://figshare.com/articles/COVID-19_Chest_X-Ray_Image_Repository/12580328].

23. COVID-CXNet [https://github.com/armiro/COVID-CXNet].

24. BIMCV-COVID-19 [https://bimcv.cipf.es/bimcv-projects/bimcv-covid19/#1590858128006-9e640421-6711].

25. Eurorad [https://eurorad.org].

26. covid-chestxray-dataset [https://github.com/ieee8023/covid-chestxray-dataset].

27. covid-19-image-repository [https://github.com/ml-workgroup/covid-19-image-repository/tree/master/png].

28. Chest X-Ray Images (Pneumonia) [https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia].

29. SIRM [https://sirm.org/category/senza-categoria/covid-19/].

30. Tuberculosis (TB) Chest X-ray Database [https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset].

31. COVID-19, SARS, MERS X-ray Images Dataset [https://www.kaggle.com/datasets/057e1b6dc41d9691e59dded4445fa8cc2f0b4b5cbcb49aef9583d95233799d5a].

32. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition: 2009*: Ieee. 2009. p. 248–55.

33. Kandel I, Castelli M. The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset. ICT express. 2020;6(4):312–5.

34. Pereira RM, Bertolini D, Teixeira LO, Silla CN Jr, Costa YMG. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Comput Methods Programs Biomed. 2020;194: 105532.

35. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: *Proceedings of the 27th European conference on Advances in Information Retrieval Research.* Santiago de Compostela, Spain: Springer-Verlag. 2005. p. 345–59.

36. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. Journal of Big Data. 2019;6(1):1–54.

37. Anand R, Mehrotra KG, Mohan CK, Ranka S. An improved algorithm for neural network classification of imbalanced training sets. IEEE Trans Neural Networks. 1993;4(6):962–9.

38. Yaqub M, Feng J, Zia MS, Arshid K, Jia K, Rehman ZU, Mehmood A. State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. Brain Sci. 2020;10(7):427.

39. Wan Z, Yuxiang Z, Gong X, Yu B. DenseNet model with RAdam optimization algorithm for cancer image classification. In: *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE): 2021*: IEEE. 2021. p. 771–5.

40. Liu C, Yin Q. Automatic diagnosis of covid-19 using a tailored transformer-like network. In: *Journal of Physics: Conference Series: 2021*: IOP Publishing. 2021. 012175.

41. An K, Zhang Y. A Self-Supervised Detail-Sensitive ViT-Based Model for COVID-19 X-ray Image Diagnosis: SDViT. Appl Sci. 2022;13(1):454.

42. Okolo GI, Katsigiannis S, Ramzan N. IEViT: An enhanced vision transformer architecture for chest X-ray image classification. Comput Methods Programs Biomed. 2022;226: 107141.

43. Dozat T. Incorporating nesterov momentum into adam. 2016.

Publisher's Note