WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Transformer Based COVID-19 Detection

Sanjay, Raja Mohammed, Surya Bharathi

Rathinam College of Arts and Science

ABSTRACT

The health crisis has been caused by the COVID19 pandemic, which has truly highlighted the necessity for timely and accurate diagnose. This paper will explore the possibility of utilizing vision transformer systems to detect COVID-19 through the means of transfer learning. It's well-known that the Vision Transformers have the capability to capture the global context aptly and decrypt the intricate visual patterns from chest X-ray images. In this work, we delved into the cutting-edge transformer architectures like ViT in order to identify COVID-19 from CXR images. Through ImageNet pretraining and transfer learning, the models managed to accomplish an impressive accuracy in the range of 78.75-89.5%. Our experiments conclusively indicate that Vision Transformers are more superior than conventional approaches, and even CNNs, in reaching the state-of-the-art performance for COVID-19 detection. The discoveries clearly postulate that Vision Transformers may be an extremely effective technique for COVID-19 diagnosis, holding significant implications for improving the efficiency and accuracy of screening and diagnosis in clinical settings.

Keywords: Vision transformers, Covid19, Chest X-rays

Introduction

The COVID-19 pandemic has resulted in tremendous global morbidity and mortality. The rapid spread of the virus placed immense strain on healthcare systems worldwide and drove up demand for accurate, reliable diagnostic tools,sometimes referred to as the coronavirus [1].The disease was first discovered in Wuhan, China,In December 2019, and spread quickly throughout the world, leading the World Health Organization (WHO) to proclaim it a pandemic in March 2020 [2]. Ever since its late 2019 inception, COVID-19 has had a significant global impact. Respiratory droplets released by an infected person when they talk, cough, or sneeze are the main way that the virus spreads [3]. Exposure to contaminated surfaces or intimate contact with an infected person can result in infection. People who have COVID-19 infection may have mild to moderate respiratory symptoms. Millions of individuals have been impacted by the illness globally, and the morbidity and fatality rates are high. The pandemic has changed daily life in numerous ways, put a burden on healthcare services, and disrupted the economy. The susceptibility of healthcare professionals and workers around the world has been made clear by COVID-19, underscoring the necessity for automated diagnosis tools to help with detection. Chest X-ray is an easily accessible, relatively affordable imaging modality that can provide critical information for diagnosing COVID-19 pneumonia, given that the virus primarily affects the lungs. However, manual interpretation of chest X-rays is time-consuming and requires expertise, which can cause dangerous delays in diagnosis and treatment, especially if spread accelerates again.

Finding those who have been infected and stopping the spread of COVID-19 have both depended on the detection of the virus. Various methods have been developed to detect the presence of the virus that causes COVID-19, SARS-CoV-2. Antigen testing, antibody testing, and polymerase chain reaction (PCR) are the three methods most frequently used to identify Covid-19 [4]. The current gold standard for COVID-19 detection is PCR. The swab (PCR-test) collects samples from either your throat or nose, which are then analyzed using a Reverse Transcription Polymerase Chain Reaction (RT-PCR) test, However, they can result in longer turnaround times and higher costs because they require specialised laboratory equipment and trained personnel.The COVID-19 pandemic has caused a great deal of morbidity and mortality around the world. The quick spread of the virus has put pressure on healthcare systems and increased demand for precise and effective diagnostic instruments. One easily accessible and reasonably priced imaging modality that can be used to identify COVID-19 pneumonia is a chest X-ray. But if it spreads quickly once more, manual interpretation of chest X-rays takes time and experience, which can delay diagnosis and treatment.

Compared to PCR testing, antigen assays are more affordable and quicker at identifying viral proteins, or antigens, in respiratory samples. They can detect those who are affected right away, providing results in a matter of minutes. But compared to PCR testing, they are less sensitive, particularly in the early stages of infection when the viral load is smaller. PCR testing may occasionally be required for confirmation because false negative results can happen. Antibodies produced by the immune system in response to an infection with SARS-CoV-2 can be detected using serological testing, commonly referred to as antibody tests[5].

These tests, which are performed on blood samples, can reveal a vaccination-induced immune response or a prior illness. Antibody assays may fail to detect antibodies in the early stages of an illness, despite their decreased utility in identifying active illnesses.Furthermore, chest X-rays and computed tomography (CT) can be used to identify COVID-19. Chest X-rays provide a rapid method of determining the extent of lung damage produced by the virus, although they are more useful in tracking the disease's development, determining the degree of lung damage, and spotting possible side effects [6]. Deep learning techniques have demonstrated potential in aiding in COVID-19 identification through the use of chest X-rays.Deep learning algorithms have been used in a number of studies to analyze chest X-rays to identify COVID19. Large datasets of chest X-rays from COVID-19 positive and negative patients were used to train these algorithms so they could identify patterns and characteristics that could differentiate between the two.In this paper, we explore the potential of ViT-based deep learning architectures for COVID-19 identification from chest X-ray images. Vision Transformers (ViTs) have demonstrated promising performance on image classification tasks, outpacing convolutional neural networks. A key advantage of ViTs is their ability to recognize long-range dependencies in images through attention mechanisms. This makes them extremely well-suited for applications like medical image analysis, where subtle patterns in images may reveal occult signs of disease. Although ViTs hold tremendous potential for medical image analysis, their application to COVID-19 detection from chest X-rays is still in early stages. Further research focused on developing and rigorously evaluating ViT-based models for COVID-19 detection is critically needed.

Related Works

Deep learning has shown immense promise in medical imaging, providing invaluable assistance in the analysis and interpretation of various medical images including ultrasound, CT, MRI, and X-ray [7]. Deep learning has catalyzed major advancements in medical imaging technologies, with burgeoning benefits and highly encouraging findings from recent research. With the rapidly growing popularity of deep learning in medical imaging, researchers have begun investigating the use of chest X-rays for COVID-19 detection.Wang et al. proposed COVID-Net, a deep convolutional neural network tailored specifically for COVID-19 detection from chest X-rays. COVIDNet relies heavily on a lightweight residual projection expansion design, enabling improved representational capacity with lower computational burden. COVIDNet achieved 93.3% accuracy across Normal, Pneumonia, and COVID-19 classes [8] Das et al. trained the Xception network for COVID-19 classification using transfer learning. Experiments on 127 COVID-19, 500 Normal, and 500 Pneumonia samples from different public datasets achieved 97% accuracy, Krishnan et al. proposed a modified vision transformer architecture for COVID detection from chest X-rays using transfer learning, attaining 97.61% accuracy. Tests were conducted on the 19,105 samples comprising the COVIDx CXR2 dataset. The 19105 samples that make up the COVIDx CXR2 dataset were used for the tests. Guefrechi et al. [10] used the COVID-19 dataset, which was created by aggregating COVID-19 and normal chest X-ray pictures from several public sources, to fine-tune three potent networks, VGG16[11]. With an accuracy of 98.30%, VGG16 fared better than the other two designs.

Syed et al. presented a novel self-supervised paradigm that uses a group-masked self-supervised framework to learn a generic representation from CXRs. The pre-trained model may then be adjusted for domain-specific tasks such as general health screening, pneumonia detection, and COVID-19. The authors obtained a 98.25% accuracy rate in representation learning using the VIT-S model[12].A bespoke transformer model that successfully distinguishes COVID-19 from normal chest X-rays with an accuracy of 98% and an AUC score of 99% was suggested by Debaditya et al. [13].

In this work, we conducted a comprehensive empirical investigation on transfer learning with transformer architectures, specifically the Vision Transformer (ViT). We leveraged the COVIDx CXR-3 dataset, currently the largest publicly available benchmark chest X-ray image dataset for computer-aided COVID-19 diagnosis. To our knowledge, this represents the first study examining the potential of different transformer topologies specifically for COVID-19 detection. We validate the transformer models using transfer learning on this dataset.

METHODOLOGY

Problem Statement

The globe has been devastated by the COVID-19 pandemic, which has resulted in widespread disease and fatalities. Preventing the spread of COVID-19 and improving patient outcomes depend on early and precise diagnosis. A common and affordable imaging modality that might offer important information for COVID-19 diagnosis is a chest X-ray (CXR). On the other hand, manual CXR picture interpretation takes a lot of time and experience, which can cause delays in diagnosis and treatment.Conventional techniques for COVID-19 identification from CXR images depend on convolutional neural networks (CNNs) or manually generated features.

We utilised the COVIDx CXR-3 dataset, which contains a sizable sample of chest X-rays from COVID-19 and pneumonia patients. This carefully curated dataset of images from multiple institutions and countries around the world serves as a standard reference dataset for chest X-ray images. The COVIDx CXR-3 dataset comprises 20,386 chest X-ray images from over 17,026 patients across at least 51 countries, making it one of the largest publicly available chest X-ray datasets. We split the X-ray dataset into a training set of 11,586 images and a held-out test set of 700 images. The training set contained 3,900 normal case X-rays and 15,994 COVID-19 case X-rays. The test set consisted of 300 COVID-19 case X-rays and 400 normal case X-rays.

ViTs are able to recognize long-range relationships in pictures since they are based on the attention process. This makes them ideal for applications like medical image analysis, where minute patterns in pictures may reveal hidden signs of illness.Although ViTs hold great potential for medical image analysis, their use in COVID-19 identification from CXR pictures is still in its infancy . Research concentrating on creating and assessing ViT-based models for COVID-19 detection is required.

PROPOSED METHOD

Datasets

The COVIDx CXR-3 dataset, which includes a sizable sample of chest X-rays from Covid19 and pneumonia patients, was used in our investigation. This dataset, which consists of pictures from patients at various institutions and nations, was assembled by skilled radiologists and doctors and serves as a standard for CXR images. As seen in Fig. 1, the COVIDx CXR-3 collection comprises 20386 chest X-ray pictures obtained from over 17,026 individuals in at least 51 countries. Of all the datasets, CXR-3 is one of the biggest. A training set of 11586 pictures and a test set of 700 images were created from the x-ray dataset. 3,900 x-rays of typical cases and 15994 x-rays of Covid19 patients made up the training set. There were 300 x-rays and 400 x-rays of Covid-19 cases in the test set

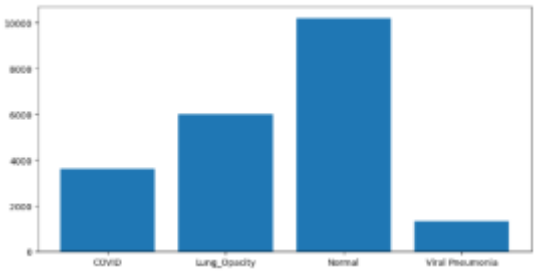

Fig. 1. COVIDx CXR-3 Visualisation

Architecture

Transformer-based architectures for COVID recognition from X-rays were investigated in this work. A deep learning architecture called Transformers is usually employed for jobs involving natural language processing[16]. Performance on picture identification tasks has increased as a result, especially on datasets containing a lot of data and intricate interactions between various image attributes. New transformer-based designs, including the Vision Transformer (ViT), have been developed as a result of the success of transformers in computer vision. These architectures have produced state-of-the-art results on a number of benchmark image recognition and object identification datasets.

Vision Transformer (ViT)

In a work called Vision Transformer (ViT), a revolutionary deep learning architecture for image identification tasks was developed. In the field of image recognition, convolutional neural networks (CNNs) have been the predominant architecture. However, ViT has presented an alternative method that processes pictures using self-attention mechanisms. A number of transformer blocks with an extra patch embedding layer make up Vision Transformers. The input picture is divided into fixed-size patches by the patch embedding layer, which then maps each patch into a high-dimensional vector representation. A trainable positional embedding is attached to each vector representation. For classification tasks, an extra trainable CLS token is linearly inserted within the patches.A feed-forward layer, a norm layer, and a multi-head self-attention layer are included in every transformer block.

While numerous attention heads aid in the learning of both local and global dependencies in a picture, the multi-head selfattention layer computes attention between a pixel and every other pixel. Prior to every multi-head self-attention module, a feed-forward layer and a normalization layer are applied. In this experiment, 16x16 fixed-size patches embedded in a linear sequence are given to the model. As seen in Fig. 2, the final output layer is changed to distinguish between Normal and Covid Chest X Rays.

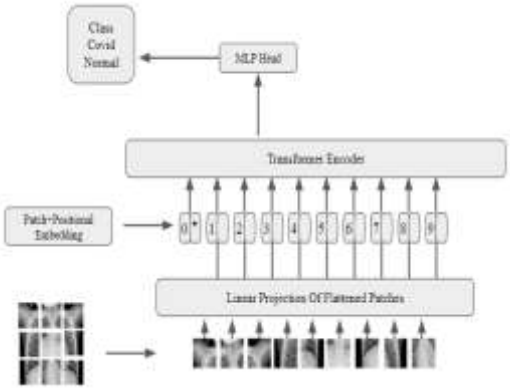

Fig. 2. Vit Architecture

Dataset Preprocessing

Using image augmentation methods , the COVID-19 photos are up-sampled. After flipping the chest X-ray either vertically or horizontally, it is randomly rotated with constant edges to a maximum of 270 degrees, and its brightness and contrast are arbitrarily altered to a maximum of 0.4.X-rays of the chest range in size from 447 × 530 to 4200 × 3290 pixels. As a result, 144 × 144 pixels was chosen as the intended picture size. is applied as an image enhancement technique because the models were pre trained on GREY photos.

The following four classes of chest pictures are included in the COVID chest x-ray datasets: Fig. 3.aCovid, Fig. 3.b.lung opacity, Fig.3.c.viral pneumonia, Fig. 3.d.Normal We had disregarded the viral pneumonia and lung opacity classes for the model's training, respectively.Normal and Covid are the two classes used to train the Vit model.

Fig. 3.a.  Covid

Fig. 3.b .  Lung Opacity

Fig. 3.c.  Viral Pneumonia

Fig. 3.d.  Normal

Fig 3 Classes In Datasets

Experiment

In the transfer learning phase, where the model is pre-trained on the ImageNet dataset, we have employed VGG[15]andVisionTransformer (ViT) models in this study. Since it only extracts generic features, the architecture's bottom layer is frozen. In order for the model to produce output from the linear layer that is particular to our dataset, we had to change the top layer.The pictures are divided into 32x32 patches for the ViT pretrained model, and the patch embeddings are obtained by passing the patches into an embedding layer. Patch embeddings are supplemented by 1-D position embeddings taken from the photos since the transformers are permutationally invariant. After that, the encoder component of ViT receives this final embedding vector.The encoder of ViTconsists of alternating Layer Normalization, Multi Headed Self-Attention and, MLP blocks. Skip connections are applied after every block.

Evaluation Criteria

Epoch and Accuracy

S.NO	Epoch	Accuracy
1	10	0.69
2	20	0.77

3	30	0.80
4	40	0.88

Fig. 4 Epoch vs Accuracy

We employed f1-score, accuracy, precision, and recall as our scoring measures in this investigation as seen in fig. 4.. Simply expressed, accuracy is the ratio of accurately predicted data to total observations, providing a fundamental understanding of the model's performance. Precision score is defined as the ratio of the accurately anticipated positive output to the total projected positive output.The Accuracy , Validation of model is plotted as shown in Fig

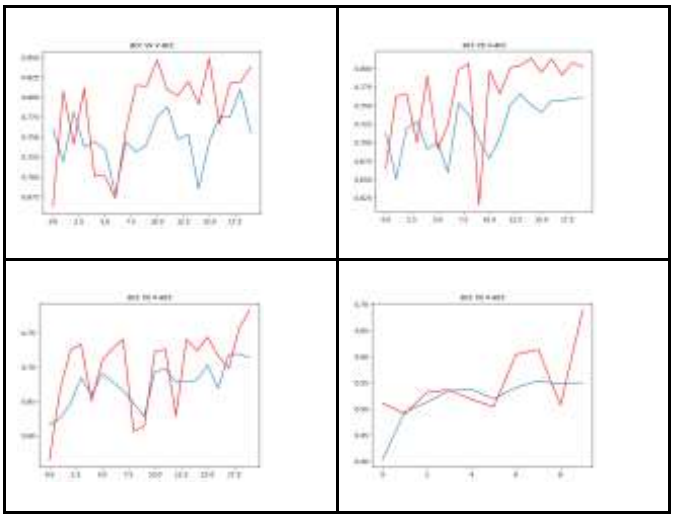

Fig.5 Accuracy vs Validation Accuracy

Using transfer learning, where the transformer networks are initialized with pretrained weights on ImageNet, we trained all the models . On the training set, we train the models, and on the test set, we report top-1 error. We employ Adam optimizer , a mini-batch with a momentum of 0.9 size of 8 and a preset weight decay of 5 times. The education rate starts at 5 x and falls as cosine increases.Every model undergoes 100 epochs of training utilizing cross-entropy loss, with five warmup epochs training, methods for enhancing data such as random cropping,Color jittering, vertical flip, and horizontal flip have all been applied to the training set at random. Python-driven models have been trained and utilized in investigations on the RTX 3090 shown in Fig.5

Results and Discussion

The ratio of the properly anticipated positive output to the total output of the real positive observation is the recall score. The F1 score, with a maximum value of 1 and a minimum value of 0, is the weighted average of the precision and recall scores.The CNN model and Vit are compared in Figures 7 and 8. it uses acc vs val_acc to evaluate the model, The blue line refers the Accuracy(acc) of the model and the red line refers the Validation Accuracy of the Model, The final Accuracy of the model is 0.88 %

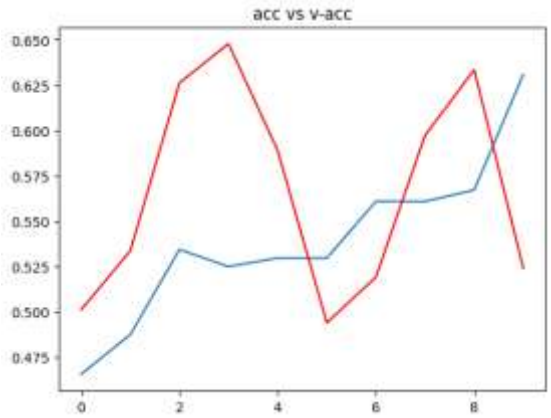

Fig. 6. CNN Model's Accuracy (blue)vs Validation Accuracy(Red)

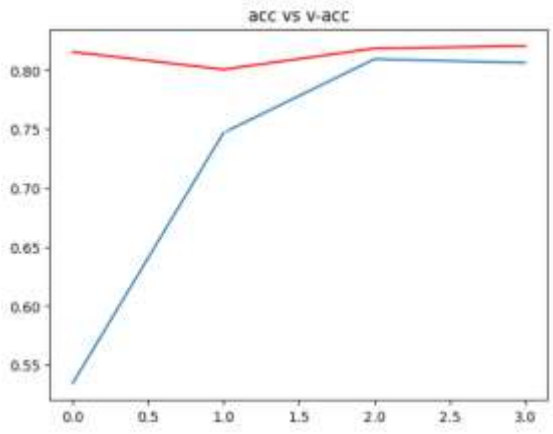

Fig. 7. ViT Model's Accuracy (Blue) vs Validation-Accuracy(Red)

Conclusion and future works:

Deep learning approaches have a lot of promise for COVID-19 diagnosis when applied to CXR image analysis. In particular,We have demonstrated that, when paired with transfer learning and the appropriate hyperparameters, transformers-based deep learning algorithms may learn long-range associations with impressive results.We are aware, nevertheless, that the study might only apply to this COVIDx CXR-3. Investigations on vision transformers' performance in 3D picture datasets and on more medical datasets are ongoing

References :

[1] C. Huang et al., 'Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China', The Lancet, vol. 395, no. 10223, 2020, doi: 10.1016/S0140-6736(20)30183-5.

[2] H. Guo, Y. Zhou, X. Liu, and J. Tan, 'The impact of the COVID-19 epidemic on the utilisation of emergency dental services', J Dent Sci, vol. 15, no. 4, 2020, doi: 10.1016/j.jds.2020.02.002.

[3] G. Rong, Y. Zheng, Y. Chen, Y. Zhang, P. Zhu, and M. Sawan, 'COVID-19 Diagnostic Methods and Detection Techniques', in Encyclopedia of Sensors and Biosensors, 2023. doi: 10.1016/b978- 0-12-822548-6.00080-7

[4]L. Wang, Z. Q. Lin, and A. Wong, 'COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images', Sci Rep, vol. 10, no. 1, 2020, doi: 10.1038/s41598-020-76550-z

[5] P. K. Chaudhary and R. B. Pachori, 'FBSED based automatic diagnosis of COVID-19 using X-ray and CT images', Comput Biol Med, vol. 134, 2021, doi: 10.1016/j.compbiomed.2021.104454

[6] A. Narin, C. Kaya, and Z. Pamuk, 'Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks', Pattern Analysis and Applications, vol. 24, no. 3, 2021, doi: 10.1007/s10044-021-00984-y.

[7] J. Feng and J. Jiang, 'Deep Learning-Based Chest CT Image Features in Diagnosis of Lung Cancer', Comput Math Methods Med, vol. 2022, 2022, doi: 10.1155/2022/4153211.

[8] I. U. Khan and N. Aslam, 'A deep-learning-based framework for

[9] Pham, T.D. A comprehensive study on classification of COVID-19 on computed tomography with pretrained convolutional neural networks. Sci. Rep. 2020, 10, 1–8.

[10] D. Shome et al., 'Covid-transformer: Interpretable covid-19 detection using vision transformer for healthcare', Int J Environ Res Public Health, vol. 18, no. 21, 2021, doi: 10.3390/ijerph182111086.

[11] A. Dosovitskiy et al., 'An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale', Oct. 2020, Accessed: Apr. 30, 2023. [Online]. Available: http://arxiv.org/abs/2010.11929

[12] Chakraborty, M.; Dhavale, S.V.; Ingole, J. Corona-Nidaan: Lightweight deep convolutional neural network for chest X-ray based COVID-19 infection detection. *Appl. Intell.* 2021, *51*, 3026–3043.

[13]automated diagnosis of COVID-19 using X-ray images', Information (Switzerland) vol. 11, no. 9, 2020, doi: 10.3390/INFO11090419.

[14] M. Z. Islam, M. M. Islam, and A. Asraf, 'A combined deep CNNLSTM network for the detection of novel coronavirus (COVID-19) using X-ray images', Inform Med Unlocked, vol. 20, 2020, doi: 10.1016/j.imu.2020.100412.

[15] S. Angara, P. Guo, Z. Xue, and S. Antani, 'An Empirical Study of Vision Transformers for Cervical Precancer Detection', in Communications in Computer and Information Science, 2022. doi: 10.1007/978-3-031-07005-1_3.

[16] O. Russakovsky et al., 'ImageNet Large Scale Visual Recognition Challenge', Sep. 2014, Accessed: 202