TYPE Original Research
PUBLISHED 27 February 2023
DOI 10.3389/fpubh.2023.1025746

CCTCOVID: COVID-19 detection from chest X-ray images using Compact Convolutional Transformers

Abdolreza Marefat[1], Mahdieh Marefat[2], Javad Hassannataj Joloudari[3], Mohammad Ali Nematollahi[4]* and Reza Lashgari[5]

[1]Department of Computer Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran, [2]Department of Cellular and Molecular Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran, [3]Department of Computer Engineering, Faculty of Engineering, University of Birjand, Birjand, Iran, [4]Department of Computer Sciences, Fasa University, Fasa, Iran, [5]Institute of Medical Science and Technology, Shahid Beheshti University, Tehran, Iran

Check for updates

OPEN ACCESS

EDITED BY
Wellington Pinheiro dos Santos,
Federal University of Pernambuco, Brazil

REVIEWED BY
Maira Santana,
Universidade de Pernambuco, Brazil
Juliana Gomes,
Federal University of Pernambuco, Brazil

*CORRESPONDENCE
Mohammad Ali Nematollahi
✉ ma.nematollahi@fasau.ac.ir

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 15 October 2022
ACCEPTED 07 February 2023
PUBLISHED 27 February 2023

CITATION
Marefat A, Marefat M, Hassannataj Joloudari J, Nematollahi MA and Lashgari R (2023) CCTCOVID: COVID-19 detection from chest X-ray images using Compact Convolutional Transformers. *Front. Public Health* 11:1025746. doi: 10.3389/fpubh.2023.1025746

COVID-19 is a novel virus that attacks the upper respiratory tract and the lungs. Its person-to-person transmissibility is considerably rapid and this has caused serious problems in approximately every facet of individuals' lives. While some infected individuals may remain completely asymptomatic, others have been frequently witnessed to have mild to severe symptoms. In addition to this, thousands of death cases around the globe indicated that detecting COVID-19 is an urgent demand in the communities. Practically, this is prominently done with the help of screening medical images such as Computed Tomography (CT) and X-ray images. However, the cumbersome clinical procedures and a large number of daily cases have imposed great challenges on medical practitioners. Deep Learning-based approaches have demonstrated a profound potential in a wide range of medical tasks. As a result, we introduce a transformer-based method for automatically detecting COVID-19 from X-ray images using Compact Convolutional Transformers (CCT). Our extensive experiments prove the efficacy of the proposed method with an accuracy of 99.22% which outperforms the previous works.

KEYWORDS

COVID-19, deep learning, Convolutional Neural Networks, vision transformers, Compact Convolutional Transformers

1. Introduction

The virus, named Severe Acute Respiratory Syndrome Corona-Virus 2 SARS-CoV-2, also known by the name COVID-19, is the source of a severe disease that started in Wuhan, China during the last months of 2019 (1). It soon spread to other parts of the globe and caused one of the most devastating pandemics, in that millions of people became abruptly affected or dead. According to the World Health Organization (WHO), the number of death cases in the first half of 2022 stood at more than 6,200,000 and the number of diagnosed people reached more than 516,000,000 in the same year worldwide. This virus belongs to the same group as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (2). Its commonly recognized symptoms are coughing, shortness of breath, fever, pneumonia, and respiratory distress (3).

The negative ramifications, imposed on the communities by this virus, and also its rapid transmission from one person to another, prove the necessity of tackling this disease with prohibitive measures. Approximately all countries included a variety of safety protocols, such as social distancing, with the object of controlling the outbreak of this pandemic. Accurately and rapidly detecting COVID-19 is an essential step that should be taken to control the widespread disease (4). Screening and monitoring of Computed Tomography (CT) and X-ray images have demonstrated great potential in providing a reliable modality for experts to examine different lung diseases such as tuberculosis, infiltration, atelectasis, pneumonia, and COVID-19 (5). However, the lack of specialized human resources in many regions, especially poor and underdeveloped countries acts as an impediment to taking advantage of such imaging technologies. This motivated the scientific community to utilize computer-aided intelligent decision-making systems to automate the required process.

Deep Learning (DL) is a powerful tool that can provide us with such systems. Among various architectures, designed for processing different types of data, Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) are specifically invented for visual data. Especially, in medical image analysis, these architectures have proven to be remarkably effective for diagnosing a wide variety of medical conditions. In the following, a brief explanation of CNNs and ViTs is given.

1.1. Convolutional Neural Network

Convolutional Neural Network (CNN) is one of the most favored types of architectures in deep learning, especially in computer vision (6). The main component of CNN-based architectures is convolution, which is a mathematical linear operation between matrices (7). CNNs' most notable success is in the field of pattern recognition applied to imagery, that is, visual data (8). In fact, the introduction of CNNs by Krizhevsky et al. (9), has revolutionized a wide variety of challenges in the domain of computer vision such as medical image analysis, face recognition, image classification, object detection, and semantic segmentation (10–15).

In general, CNN-based models comprise three types of layers, namely convolutional layers, pooling layers, and fully-connected layers. These three are depicted in Figure 1, where you can see a formation of a CNN-based model for classifying the input lung X-ray image into healthy or unhealthy samples. As is shown in this figure, the convolution layer operates by sliding a kernel on the input data. Each kernel results in a feature map, to which the pooling operation is applied.

Furthermore, translation equivariance and translational invariance, which is inherent to CNNs, enable them to learn the natural statistics of the input image. In addition to this, sparse interaction, weight sharing, and equivariant representations make CNN-based models more efficient and less computationally expensive (16).

1.2. Vision transformer

Transformer-based models in deep neural networks have been originally introduced in the domain of Natural Language Processing (NLP) (17). The astounding performance of these models in a variety of tasks in NLP, i.e., machine translation (18), question answering (19), text classification (20), and sentiment analysis (20, 21), has sparked the interest of a considerable number of researchers in computer vision to attune these models to the field of computer vision (22, 23) was the first research paper, in which the authors creatively invented a way to apply transformers to the visual data and introduced ViTs for image classification. Figure 2 demonstrates a general procedure in ViT-based models. Based on this figure, it can be witnessed that an image is converted to a set of patches, each representing a locality of a region in the image. This procedure enables us to look upon an image as sequential data; the type of data that is prevalent in NLP and is tailored for transformers.

Firstly, ViT flattens these patches and then passes them through a trainable linear projection layer, making the projections the same with regard to their dimensionality. Then, since the ViT is thoroughly agnostic to the hierarchy of the input image, meaning that it does not take into consideration where each patch is located in the original image, the position embeddings are integrated into these projections to eradicate this problem. After that, the transformer encoder block takes these patches, alongside their position, and an extra classification token named CLS token. The transformer encoder includes multi-head attention layers, capable of learning a variety of self-attention states. Lastly, the outputs of all existent heads are amalgamated and fed to the Multi-Layer Perceptron (MLP).

2. Related works

In this section, we present a brief review of the previous works for detecting COVID-19 from CT or X-Ray images. Due to the successful performance of deep learning-based approaches in medical image analysis (24), researchers have focused on proposing different CNN or ViT-based architectures in order to automatically recognize the presence of the infection (5).

To begin with, Wang et al. (25) were one of the first groups who designed a deep neural network for detecting COVID-19. In addition to this, they provided a relatively large dataset of chest X-ray images. They achieved 93.3% accuracy. In (26), Marques et al. proposed a pipeline based on EfficientNet and followed the 10-fold cross-validation paradigm to evaluate their approach to chest x-ray images. They have achieved an average accuracy of 99.62 and 97.54% in binary and multi-class classification, respectively. Singh et al. (27) utilized a famous neural network, named VGG16, and transfer learning in order to detect COVID-19 from CT scans. In their approach, the extracted features were chosen by using Principal Component Analysis (PCA) and later classified by different classifiers. At most, they achieved 95.7% accuracy. In (28), Islam et al. made a neural network that was a hybrid of CNNs and Long Short-Term Memory (LSTM) networks. They trained their model on 3 classes, namely COVID-19, pneumonia, and normal, and achieved 99.2, 99.2, and 99.8% accuracy for

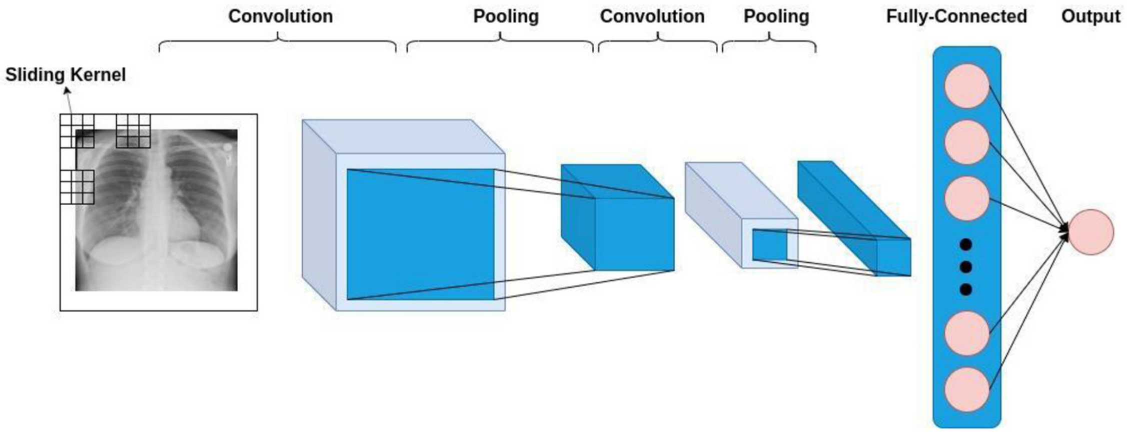

FIGURE 1
The general architecture of a CNN-based model.

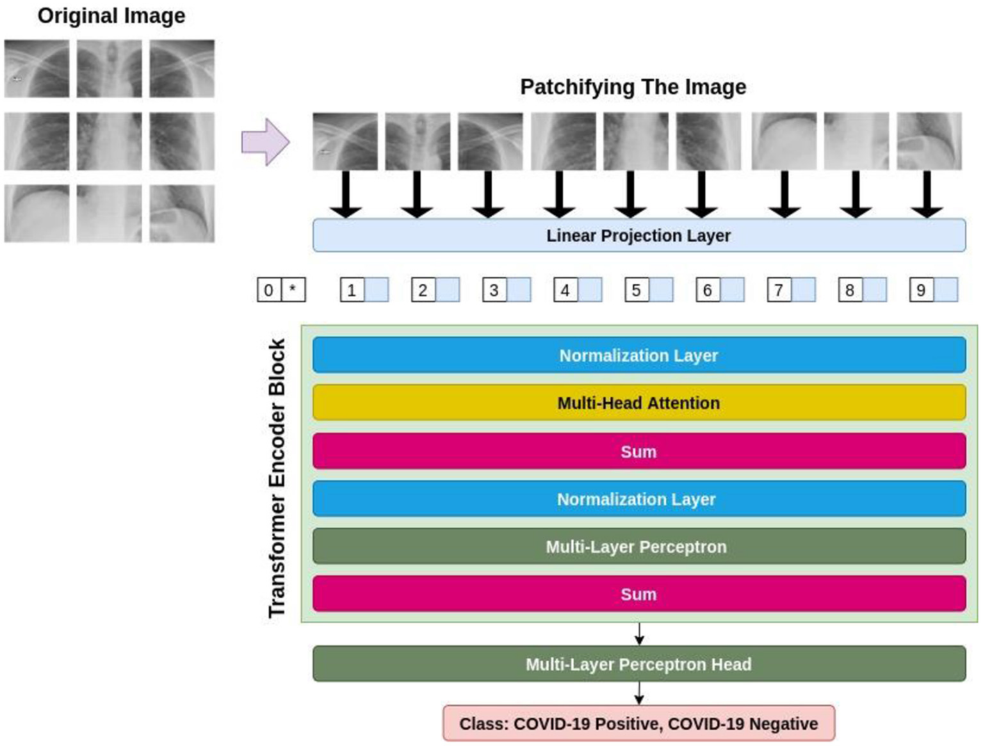

FIGURE 2
The general architecture of a ViT-based model.

each class, respectively. In Narin et al. (29), have thoroughly investigated the impact of transfer learning on the analysis of chest X-ray radiographs. Five pre-trained models, namely ResNet50, Resnet101, ResNet152, InceptionV3, and Inception-ResNetV2 were the models examined by them and they achieved accuracies of 96.1, 99.5, 99.7% in three different datasets. In addition, Goel et al. (30) have proposed OptCoNet; an optimized Convolutional Neural Network for detecting COVID-19 from X-ray images. They employed the gray wolf optimization algorithm with the aim of tuning the hyperparameters of the classifier and achieved 97.78% accuracy.

Furthermore, more recently ViT-based models have been put forward for COVID-19 detection. In Al Rahhal et al. (31), a novel model with two branches has been proposed. In this work, a ViT architecture is utilized as a backbone, integrated with a Siamese network for processing an augmented version of the input X-ray image. They could obtain 99.13% for their accuracy in the 80:20 distribution of train and test. Further, Mondal et al.

TABLE 1 An overview of the existing works.

References	Dataset (CT/X-ray)	Classification	Approach	Train/test/validation	Performance (accuracy)
Konar et al. (35)	X-Ray	Binary	Proposed Semi-Supervised Classifier	Random sampling 70% for training and 30% for testing	98.40%
Vaid et al. (36)	X-Ray	Binary	VGG-19	Random sampling 80:20:20 for train, validation and testing.	96.30%
Ozturk et al. (37)	X-Ray	Binary and multi-class	DarkNet	5-fold cross-validation	98.08%
Panwar et al. (38)	X-Ray	Binary	Proposed nCOVnet	Random sampling 70% for training and 30% for testing	97.62%
Ahuja et al. (39)	X-Ray	Binary	ResNet-18	Random sampling. 70% for training and 30% for testing.	99.40%
Sharifrazi et al. (40)	X-Ray	Binary	Sobel+Support Vector Machine +CNN	10-fold cross-validation	99.02%
Khozeimeh et al. (41)	Numerical	Binary	CNN-AE	10-fold cross-validation	96.05%
Al Rahhal et al. (31)	CT/X-Ray	Multi-class	Proposed Siamese+ViT Classifier	60:40 80:20 20:80	99.13 ± 0.23
Mondal et al. (32)	CT/X-Ray	Multi-class	Proposed xViTCOS + Multistage Transfer Learning	80:20	0.981
Krishnan and Krishnan (2)	X-Ray	Binary	Pretrained ViT	73:3:24	97.61
Kumar et al. (42)	X-Ray	Multi-class	SARS-Net CNN	90:10	97.60
Esmi et al. (43)	X-Ray	Multi-class	Fuzzy fine-tuned Xception	80:20	96.60

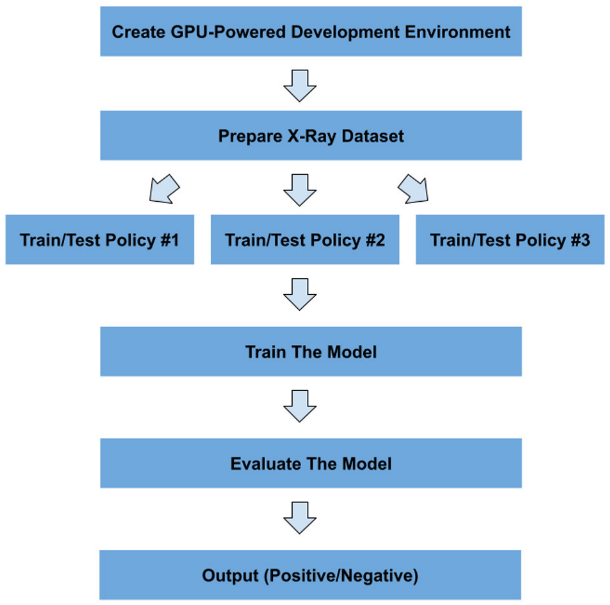

FIGURE 3
Workflow of the proposed pipeline for detecting COVID-19.

TABLE 2 The dataset distribution.

	No. train samples	No. test samples
Positive (COVID-19)	16,490	200
Negative (NO COVID-19)	13,992	200
Total	30,482	400

(32) proposed a ViT-based model and employed a multi-stage transfer learning technique to address the scarcity of data. They obtained an overall accuracy of 96.00%. Furthermore, Liu et al. (33) have applied a transformer-like classifier model. By employing transfer learning techniques in their approach, they improved the results to outperform CNN-based models, achieving 99.7% accuracy. Additionally, in (34), the authors applied a ViT-based algorithm based on the Swin transformer for feature learning and aggregation in two stages segmentation and classification. In their paper, they further validated the superiority of their algorithm by comparing their results with well-known visual feature extractors, i.e., EfficientNetV2. The accuracy of 94.3% was obtained by their approach.

Furthermore, we have provided Table 1, which details an overview of the existing research works on the diagnosis of COVID-19 from CT or X-ray images.

In contrast to the efficiency of previous works, the related literature lacks employ ViT-based deep models with less hunger for data. Although such models lack inductive biases like translation equivariance and locality, which are inherent to CNN-based models, they are not efficient in generalizing in the procedure of training on small datasets and this shows their data-driven nature which is not feasible, especially in the medical area, where it is less likely to have access to the huge amount of data. As a result, ViTs do not seem to be a better choice when dealing with

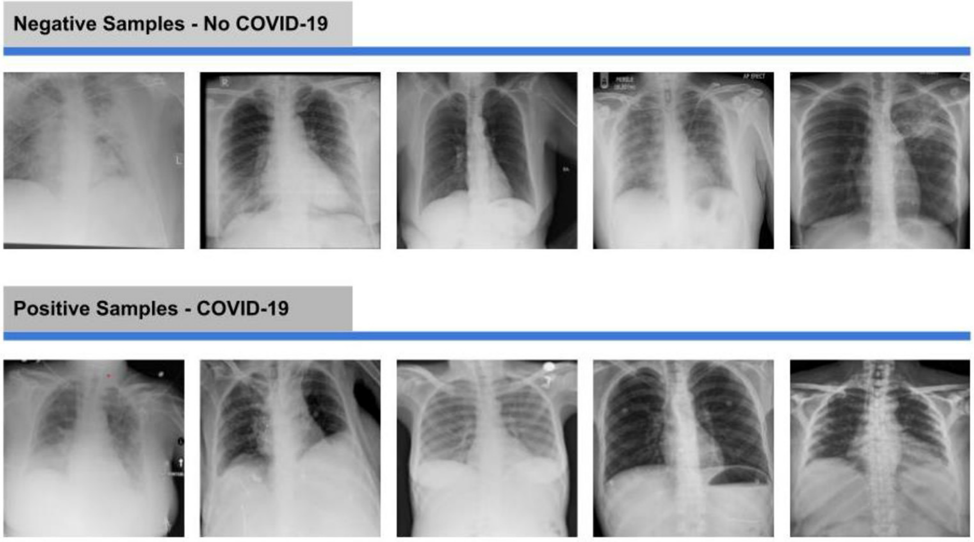

FIGURE 4
Negative and positive samples of X-ray images from the dataset.

small datasets because they have more requirements both in terms of computation and memory, preventing many researchers from adopting such models in different areas. The above-mentioned challenges motivated us to propose a more performant solution that utilizes both the CNN and ViT-based models simultaneously with the object of boosting COVID-19 detection from visual data.

3. Methodology

This section includes our methodology for detecting COVID-19 from X-ray images. The workflow of the adopted pipeline is shown in Figure 3.

Moreover, in this section, after describing the details of the used dataset, all the main components of the proposed method will be elaborated.

3.1. Dataset description

In this paper, a publicly available dataset[1] is used for training and evaluating our proposed method. Table 2 shows the official distribution of this dataset.

Moreover, Figure 4 demonstrates some samples from both positive and negative classes.

3.2. The used architecture

This section introduces our proposed algorithm, including different stages in Compact Convolutional Transformers (CCT) (16). The overview of CCT architecture is illustrated in Figure 5.

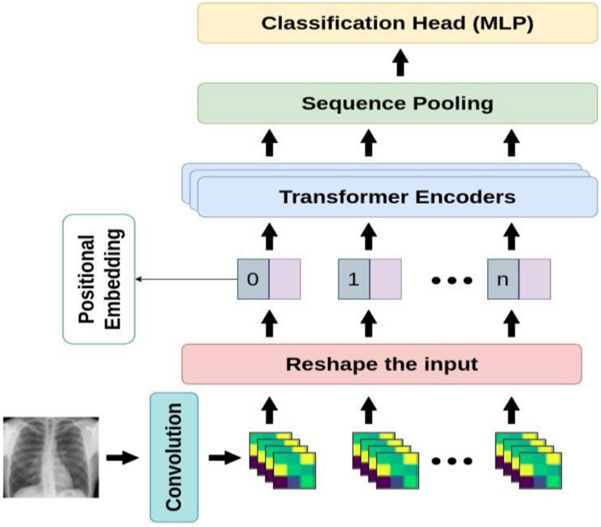

FIGURE 5
An overview of the proposed architecture.

Compact Convolutional Transformer (CCT) is one of the most recent compact transformer-based models for image processing. The biggest advantage of CCT is its ability to learn from the small amount of data in comparison with the base ViT models which are considerably data-hungry. In situations where we do not have access to huge number of samples, ViT models cannot achieve better performance than convolutional-based models and thus such models are not considered suitable. However, although some state-of-the-art models, namely DeiT, ConViT, and Compact Vision Transformers have addressed the issue of the low volume of data in ViTs, CCT has managed to outperform all other pre-existing approaches. CCT's performance has been studied through

[1] https://www.kaggle.com/datasets/andyczhao/covidx-cxr2?select=competition_test

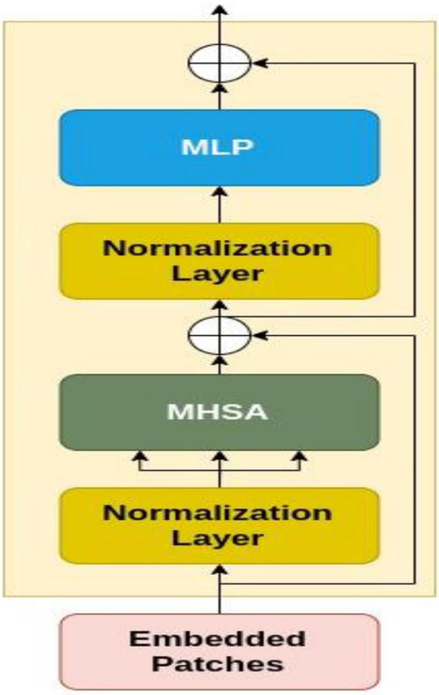

FIGURE 6
The architecture of transformer encoder.

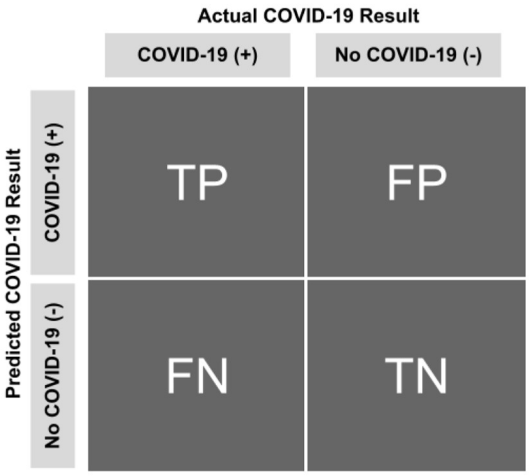

FIGURE 7
Confusion matrix.

a variety of low to high-resolution images in benchmark datasets such as FashionMNIST, MNIST, CIFAR-10, CIFAR-100, ImageNet, and Flowers-102.

Moreover, CCT is developed over Compact Vision Transformers (CVT) and takes advantage of a convolutional tokenizer leading to the preservation of local information and the production of richer tokens. Compared to the primitive ViT, the convolutional tokenizer is more effective in encoding the connection between patches. In the sequel, we go into further detail on the components of compact transformers.

TABLE 3 The metrics used for evaluation.

Metric name	Equation
Accuracy	$\frac{TP+TN}{FP+FN+TP+TN}$ (2)
Precision	$\frac{TP}{TP+FP}$ (3)
Recall	$\frac{TP}{TP+FN}$ (4)
F1-score	$\frac{(2*precision*recall)}{precision+recall}$ (5)
AUC-ROC	Area under curve of receiver operator characteristic (6)
False positive rate (FPR)	FP/(FP + TN) (7)
False negative rate (FNR)	FN/(TP + FN) (8)
True negative rate (TNR)	TN/(TN + FP) (9)

TABLE 4 Experimental setup.

Programming language	Python 3.7
Deep learning library	Pytorch 1.9
CPU	Intel® Core^{TM} i7-10700 CPU @ 2.90 GHz × 16
GPU	GeForce GTX 1060

TABLE 5 Hyperparameter settings.

Parameter name	Detail
Image size	(256, 256)
Embedding dimension	512
Number of convolution layers	4
Pooling kernel size	5
Pooling padding	1
Pooling stride	2
Kernel size	5
Stride	2
Padding	1
Number of heads	8
Number of classes	2
Positional embedding	Sine function

3.2.1. Transformer-based backbone

As for CCT model design, the original Vision Transformer (23), and original Transformer (44) are proposed. The encoder is made up of transformer blocks, each of which has an MLP block and a Multi-Head Self-Attention (MHSA) layer. Based on Figure 6, the input image is patchified, after which each patch becomes flattened and projected linearly. Then, the positional embeddings are added to these patch embeddings. These embeddings are fed to multiple transformer encoders, whose architecture is shown in Figure 6 in detail.

Figure 6 demonstrates that the embedded patches are fed to a normalization layer and after that MHSA module is applied to

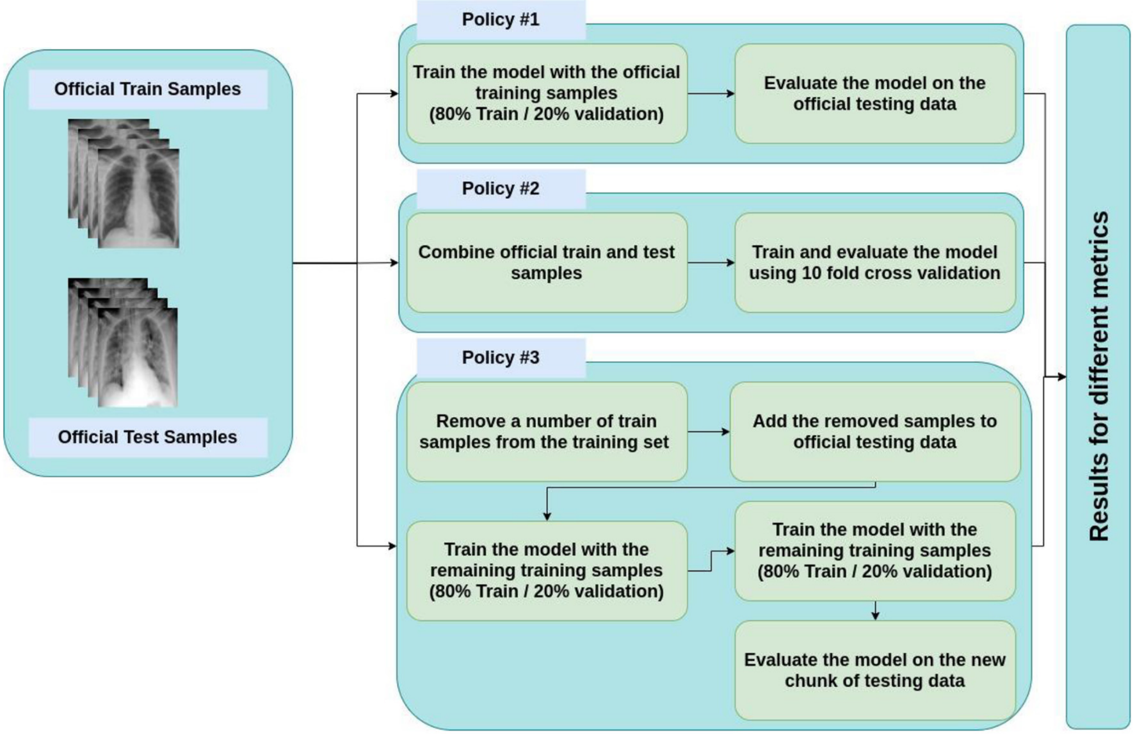
Official Train Samples
Official Test Samples
Policy #1
Train the model with the official training samples (80% Train / 20% validation)
Evaluate the model on the official testing data
Policy #2
Combine official train and test samples
Train and evaluate the model using 10 fold cross validation
Policy #3
Remove a number of train samples from the training set
Add the removed samples to official testing data
Train the model with the remaining training samples (80% Train / 20% validation)
Train the model with the remaining training samples (80% Train / 20% validation)
Evaluate the model on the new chunk of testing data
Results for different metrics

FIGURE 8
Policies for train and test split.

TABLE 6 Results of classification on the official test data (all metrics are reported on a 0–100 scale).

Accuracy	Precision	Recall	F1-score	AUC-ROC	FPR	FNR	TNR
99.00	99.00	99.00	99.00	99.67	1.00	1.00	99.00

the output. Also, the residual connections exist before and after each block of MLP. The encoder additionally employs dropout, GELU activation, and Layer Normalization. It is considered that the vision transformers are more compact and simpler. Variants with (the minimum number of) 2 layers, 2 heads, and 128-dimensional hidden layers are implemented. Based on the image resolution of the training dataset, the tokenizers are modified. These variations are referred to as ViT-Lite, and although they differ in size, they are largely comparable to ViT in terms of architecture.

3.2.2. Sequence pooling

The ViT and almost all general transformer-based classifiers follow BERT (45), which sends a learnable class or query token across the network before feeding it to the classifier leading to the conversion of the sequential outputs to a single class index. However, in CCT, an attention-based technique that pools over the output token sequence are leveraged, and hence, unlike the learnable token, the output sequence contains substantial information that includes several parts of the input image, resulting in a more efficient performance. Moreover, the network can correlate data across the input data and weigh the sequential embedding of the transformer encoder's latent space. Finally, Compact Vision Transformer (CVT) is made by substituting SeqPool for the ordinary class token in ViT-Lite.

3.2.3. Convolutional tokenizer

As for the last steps in designing CCT, a straightforward convolutional block is substituted for the patch and embedding in ViT-Lite and CVT to induce an inductive bias into the model. A single convolution, ReLU activation, and a max pool make up the standard and customary design of this block by which the models have more flexibility than models like ViT since they are no longer restricted to input resolutions that are strictly divisible by the predetermined patch size. The CCT is produced *via* this convolutional tokenizer, whose mathematical representation is shown in Equation (1), Sequence Pooling, and the transformer encoder.

$$X_0 = MaxPool(ReLU\left(conv2d(x)\right)) \tag{1}$$

The feature map is extracted to be the representation of local features. Based on Equation 1, we can deduce that CCT does not depend on image resolution, since it preserves locality in information gained from the data due to its convolutional blocks.

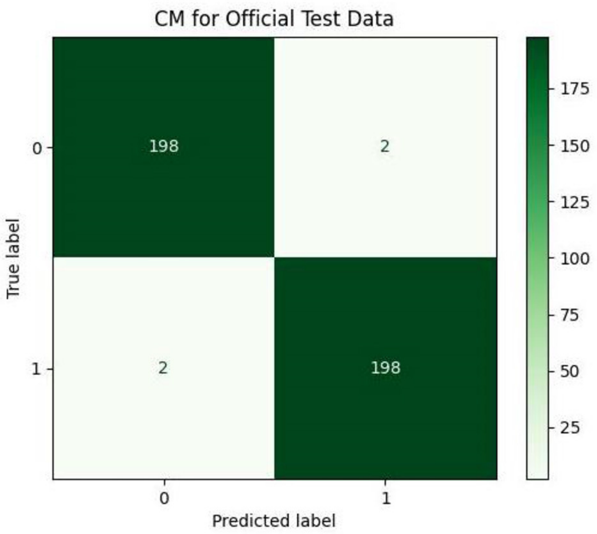

FIGURE 9
Confusion matrix (CM) for official test data.

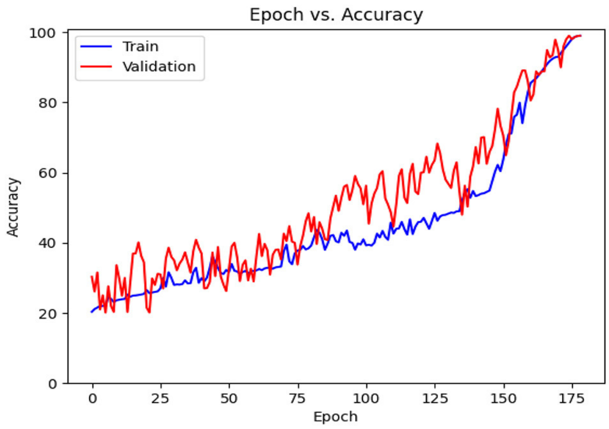

FIGURE 10
Accuracy diagram vs. epochs.

3.3. Evaluation metrics

The measures used for evaluating the performance of the proposed classifier are estimated against the following metrics:

Confusion Matrix (CM): A matrix, containing four main elements, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). For a binary classifier, CM is as Figure 7.

True Positive (TP): the number of infected samples correctly classified as infected.

True Negative (TN): the number of uninfected samples correctly classified as uninfected.

False Positive (FP): the number of infected samples correctly classified as infected.

False Negative (FN): the number of infected samples correctly classified as infected.

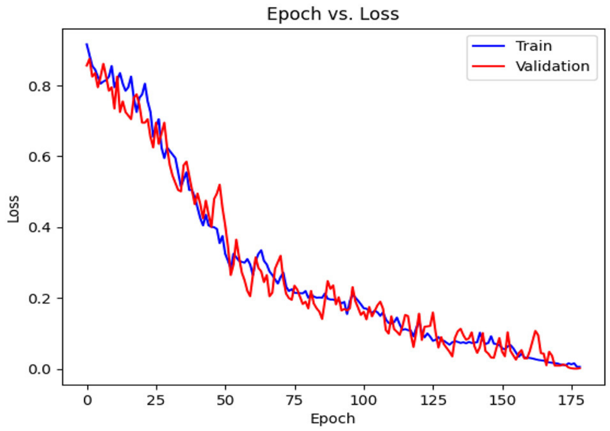

FIGURE 11
Loss diagram vs. epochs.

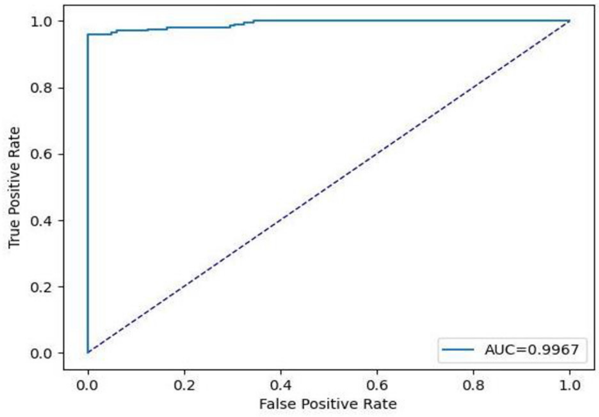

FIGURE 12
Receiver operating characteristic (ROC) curve for official test data.

Based on the metrics mentioned above, the metrics detailed in Table 3 can be deduced and used for evaluating a classifier. Other metrics used for evaluating the proposed approach are accuracy, precision, recall, F1-Score, AUC-ROC, False Positive Rate (FPR), False Negative Rate (FNR), and True Negative Rate (TNR) (40, 46–48).

4. Results

This section includes the results of classification by our proposed approach.

4.1. Experimental setup

Table 4 details the software and hardware used for implementing our proposed method.

TABLE 7 Results of classification using 10-fold cross-validation (all metrics are reported in 0–100 scale).

No. fold	Accuracy	Precision	Recall	F1-score	AUC-ROC	FPR	FNR	TNR
1	99.16	98.90	99.42	99.16	99.23	1.10	0.58	98.90
2	98.90	98.21	99.61	98.91	99.15	1.81	0.39	98.19
3	99.71	99.68	99.74	99.71	99.20	0.32	0.26	99.68
4	99.38	99.10	99.68	99.39	99.61	0.91	0.32	99.09
5	98.96	98.40	99.55	98.97	99.03	1.62	0.45	98.38
6	99.61	99.42	99.81	99.61	99.27	0.58	0.19	99.42
7	99.03	98.84	99.22	99.03	99.19	1.17	0.78	98.83
8	98.80	97.90	99.74	98.81	99.02	2.14	0.26	97.86
9	99.64	99.68	99.61	99.64	99.64	0.32	0.39	99.68
10	99.03	98.65	99.42	99.03	99.31	1.36	0.58	98.64
Average	**99.22**	**98.88**	**99.58**	**99.23**	**99.27**	**1.13**	**0.42**	**98.87**

The bold values demonstrate that the best values is placed in the last record.

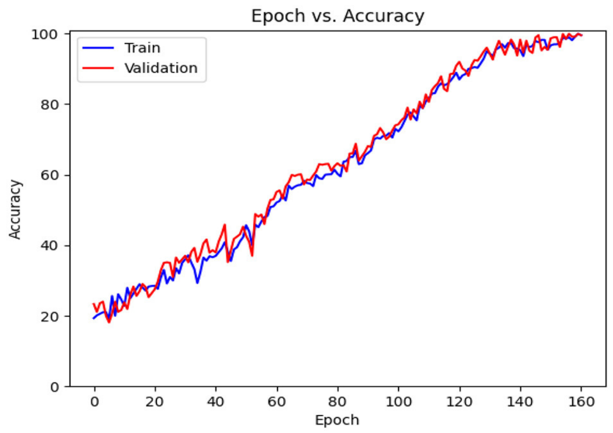

FIGURE 13
Accuracy diagram vs. epochs.

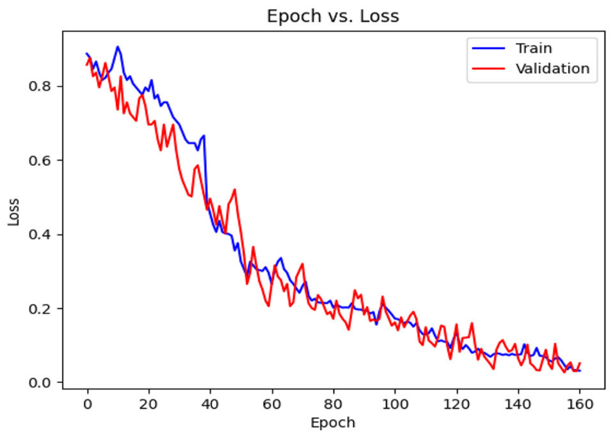

FIGURE 14
Loss diagram vs. epochs.

4.2. Hyperparameter settings

Table 5 details the Hyperparameter settings applied for implementing our proposed method.

4.3. Dataset split

Note that we opted for three main policies for evaluating the classifier. These three are:

1) Policy #1: We used the official training data for training and validating the model and the official test data for testing it.
2) Policy #2: We amalgamated official train and test data with each other; then randomly shuffled the data multiple times. Next, we used 10-fold cross-validation method for the training and evaluation process.
3) Policy #3: We randomly shuffled the training data multiple times and then chose a specific number of training data (randomly chosen), removed them from the training set, and added them to the testing set. The number of replaced samples was set in a way to make the test size 0.1 of the remaining training data.

The main reason for pursuing these policies is the small size of the official test chunk, which makes the evaluation results unreliable. This process is depicted in Figure 8.

4.3.1. Results for policy #1

Table 6 contains the results of classification by the proposed model on the official distribution of the used dataset. Additionally, Figure 9 shows the obtained CM for the same distribution. Figures 10, 11 show the accuracy and loss curves vs. epochs, respectively. Figure 12 demonstrates the ROC curve for the classifier.

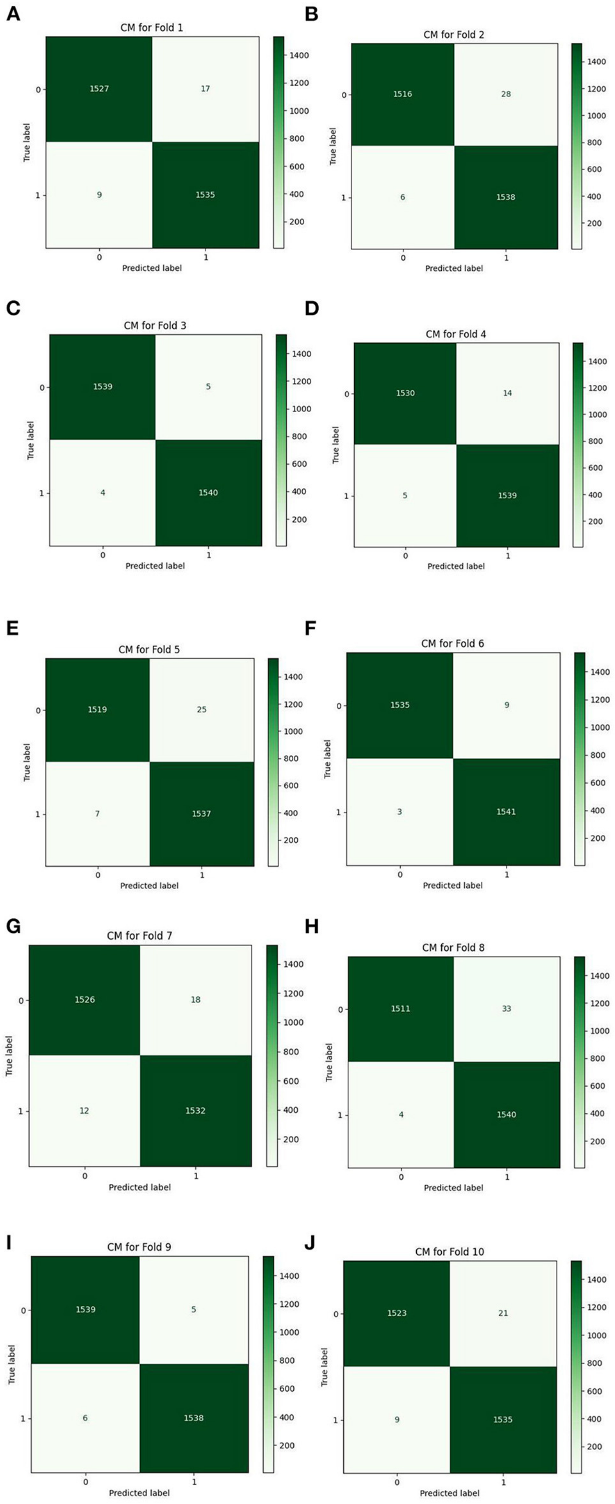

FIGURE 15
The CMs based on 10-folds **(A–J)**.

TABLE 8 Train and test distribution in policy #3.

	No. train samples	No. test samples
Positive (COVID-19)	26142	4740

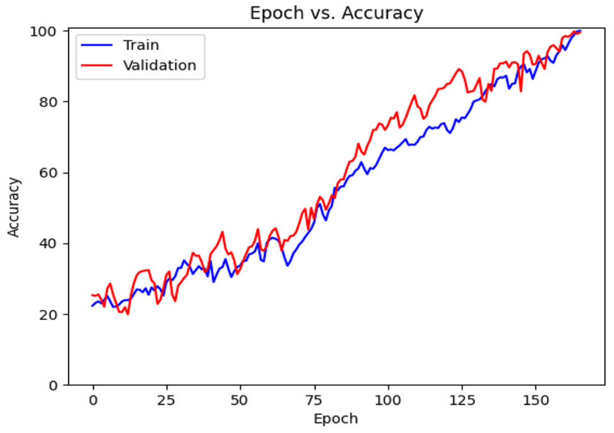

FIGURE 16
Accuracy diagram vs. epochs.

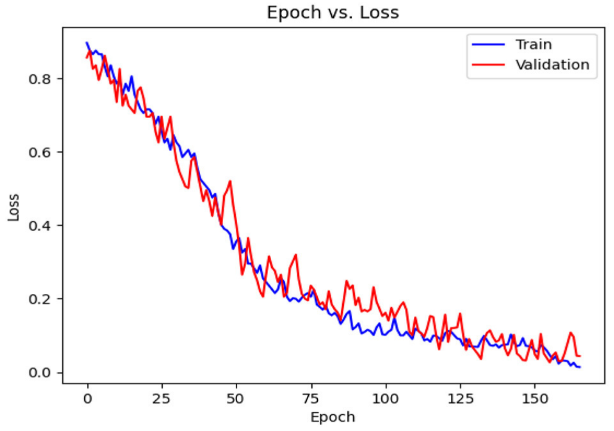

FIGURE 17
Loss diagram vs. epochs.

Based on Table 6, it can be observed that our approach can achieve a high value of 99.00% for accuracy, precision, recall, and F1-Score. The stability of the proposed model in terms of detecting both negative and positive samples can be proved by the fact that a high value of 99.67 is achieved for AUC-ROC.

4.3.2. Results for policy #2

Table 7 details the achieved results for each fold based on the introduced metrics. Figures 13, 14 demonstrate the accuracy and validation curves achieved in the training process. Figures 15A–J shows the CMs obtained in the second policy.

Table 7 shows the results achieved in all folds as well as the average. The achieved accuracy, on average, is 99.22, the precision is 98.88, the recall is 99.58, and the F1-Score is 99.23. The value for AUC-ROC, on average, is 99.27 which shows the strong confidence

TABLE 9 Results of classification using policy three (all metrics are reported on a 0–100 scale).

Accuracy	Precision	Recall	F1-score	AUC-ROC	FPR	FNR	TNR
99.09	98.74	99.45	99.09	99.73	1.27	0.55	98.73

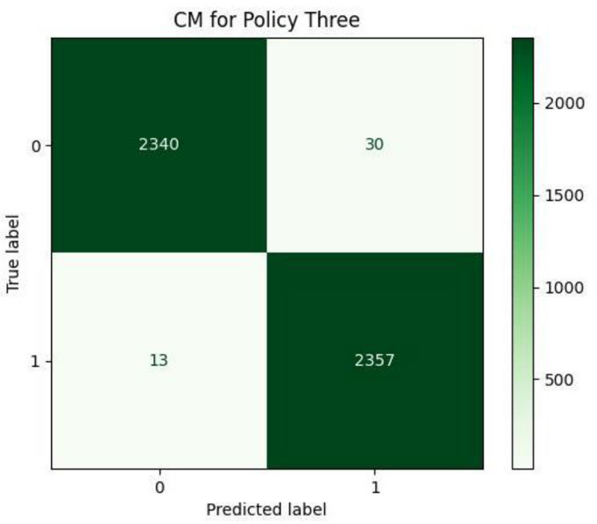

FIGURE 18
Confusion matrix.

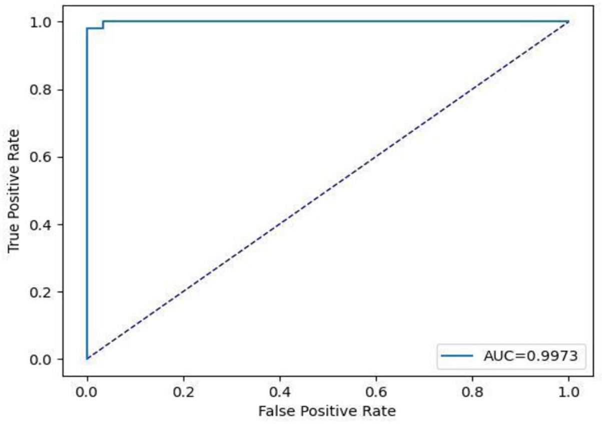

FIGURE 19
Receiver operating characteristic curve.

of the proposed classifier in classifying both negative and positive samples.

4.3.3. Results for policy #3

This subsection includes our results based on the third evaluation policy. Train and test distribution in policy #3 is indicated in Table 8. Figure 16 demonstrates training and validation accuracy in each epoch. Also, Figure 17 illustrates training and validation loss in the training procedure. Table 9 shows the results achieved by the classifier when we adopt policy 3 for the evaluation. Also, the obtained CM and ROC, in this policy, is shown in Figures 18, 19, respectively.

Based on Table 9, we can see that in policy 3, we have achieved 99.09 as accuracy, 98.74 as precision, 99.45 as recall, and 99.09 as F1-Score. 99.73 is achieved as the AUC-ROC of the classifier in policy 3 and proves the efficient performance of the model in distinguishing positive and negative samples correctly.

5. Discussion

The COVID-19 serious illness that began in the final months of 2019 and quickly spread to other regions of the world, led to one of the most destructive pandemics. The WHO estimates that as of August 2022, there have been more than 6.4 million deaths and 570 million confirmed cases. According to the research and experiences obtained up to now, CT scans and X-ray images are highly effective tools in diagnosing COVID-19. The absence of specialized human resources in many areas makes it difficult to benefit from such imaging technologies that are encouraged. The scientific community uses computer-aided intelligent systems to automate the desired procedure.

In this study, we proposed deep learning methods for the detection of COVID-19, based on X-ray images of both confirmed (positive) and negative COVID-19 cases that were gathered in a dataset with 30,882 samples. The main architecture that we proposed was CCT. Because of its compactness, CCT can be implemented in low-resource environments, which is its primary advantage, and therefore, is considered to be among mobile-friendly models. In addition, because CCT is a hybrid model based on CNN and ViT, it combines the benefits of both while avoiding their drawbacks. For instance, CCT experiences substantial performance improvements, resulting in a top-1% accuracy of 98% on CIFAR-10.

The CCT model is the only transformer-based model among the top 25 models in terms of performance and efficiency on CIFAR-10, despite having no pre-training and being rather small compared to the majority. Moreover, CCT surpasses the majority of comparable CNN-based models in this field, except for some Neural Architectural Search techniques (49). Furthermore, CCT can be lightweight, using only 0:28 million parameters, while still achieving accuracy within 1% of the top 90% of similar models on CIFAR-10. CCT obtains 80.67% accuracy on ImageNet with fewer parameters and less computational work, and it outperforms more recent, comparably sized models like DeiT (50) [for more information, see (16)].

In order to achieve better performance in our study, we chose to evaluate the classifier according to three primary policies. In policy 1, we merely trained and validated the model through the official training data, and we examined the classifier using the official test data. Afterward, to obtain more reliable and robust outcomes, the

TABLE 10 Comparison between the current study and related studies based on binary classification for COVID-19 diagnosis.

References	Cases/normal/COVID	Method	Performance (accuracy, precision, recall, F1-score, AUC, NC: Not considered)
Alakus and Turkoglu (51)	600/520/80	CNN-LSTM	Accuracy: 92.30% Precision: 92.35% Recall:93.68% F1-Score: 93.00% AUC:90.00%
Oguz and Yaganoglu (52)	1,345/738/607	ResNet-50+SVM	Accuracy: 96.29% Precision: 96.66% Recall: 95.08% F1-Score: 95.86% AUC:98.21%
Srivastava et al. (53)	4,551/3,270/1,281	lightweight CNN (CoviXNet)	Accuracy: 99.56% Precision:100% Recall: 99.70% F1-Score: 100% AUC: 99.00%
Chen (54)	60,000/59,600/400	CNN + Histogram-oriented gradients	Accuracy: 92.95% Precision: 91.5% Recall: 85% F1-Score: N/C AUC: N/C
Nasiri and Hasani (55)	625/500/125	DenseNet-169 + XGBoost	Accuracy: 98.23% Precision: 98.54% Recall: 92.08% F1-Score: 97.00% Specificity: 99.78% AUC: N/C
Jain et al. (56)	1,832/1,372/460	ResNet-101	Accuracy: 98.93% Precision: 96.39% Recall:98.93% F1-Score: 98.15% AUC: 98.20%
Al-Waisy et al. (57)	800/400/400	A hybrid deep learning detection System (deep belief network + convolutional deep belief network)	Accuracy: 99.93% Precision: 100% Recall: 99.90% F1-Score:99.93% AUC: N/C
Ardakani et al. (58)	1,020/510/510	ResNet-101	Accuracy: 99.51% Precision: 99.03% Recall: 100% F1-Score: N/C AUC: 99.40%
Zhao et al. (59)	16,351/14,704/1,647	Big transfer-M	Accuracy: 96.50% Precision:100% Recall: 93.00% F1-Score: N/C AUC: 99.40%
Haghanifar et al. (60)	3,628/3,200/428	A 121-layer dense Convolutional network	Accuracy: 99.04% Precision: N/C Recall: N/C F1-Score: 96.00% AUC: N/C
In this study (2023)	30,882/14,192/16,690	Compact Convolutional Transformers	Accuracy: 99.22% Precision:98.88% Recall: 99.58% F1-Score: 99.23% AUC: 99.27%

official test and train data were combined, after which they were repeatedly shuffled at random. The training and evaluation process was then conducted using the 10-fold cross-validation method which altogether constitutes our second policy. Finally, to provide the third (and the last) policy, we shuffled the training data at random several times followed by randomly selecting a group of training data, taking them out of the train set, and adding them to the testing set. It is important to note that the official test chunk's small size, which renders the evaluation results untrustworthy, was our main motivation for considering these three different policies and approaches.

Table 10 lists the comparison between the current study and several related studies on the topic of COVID-19 diagnosis based on binary classification, and the performance of each study is mentioned due to the evaluation metrics used by the authors.

Following is a brief description of the methodology and results of the articles listed in the table above. In Alakus and Turkoglu's study (51), six different deep-learning model types were developed and the outcomes were compared. With an accuracy of 92.30%, CNN-LSTM produced the best results out of the group.

In (52), 1,345 CT scans were subjected to deep feature extraction using deep learning models like ResNet-50, ResNet-101, AlexNet, etc. Following that, classification methods were given the deep features, and test images were used for model evaluation. The results showed that ResNet-50 and the SVM together provided the best performance. The F1-score was 95.86%, the accuracy was 96.29%, and the AUC was 98.21%.

Srivastava et al. in (53) proposed CoviXNet, a lightweight CNN-based model, over a dataset of three classes: COVID-19, normal X-rays, and viral-pneumonia-infected chest X-ray images, with an accuracy of 99.56% for binary classification (i.e., normal Chest X-ray image and COVID-19 infected).

The literature study (54) suggested a CNN-based plus histogram-oriented gradients (HOG) model on a public dataset of 60,000 X-ray images with 59,600 negative and 400 positive COVID-19 samples and a 92.95% accuracy was attained.

In (55), features from 1,125 X-ray images, including 125 images identified as COVID-19 were extracted using DenseNet-169. The XGBoost classifier was then fed the derived features and the average accuracy was 98.23%.

A deep learning ResNet-50 network was utilized as a classifier in the study (56) to identify viral/bacterial pneumonia and normal cases among 1,832 X-ray chest images. Additionally, the ResNet-101 was employed to determine COVID-19 in patients with positive viral-induced pneumonia and the overall accuracy was 98.93%.

A parallel design (COVID-DeepNet) that combines a deep belief network with a convolutional deep belief network trained from scratch on a large dataset was proposed by Al-Waisy et al. (57). With a 99.93% detection accuracy rate, the method properly identified COVID-19 in patients.

Ten well-known deep learning-based techniques for distinguishing COVID-19 from non-COVID-19 in CT scan images were proposed by Ardakani et al. (58), and the results showed that the ResNet-101 model achieved 99.51% accuracy.

To detect COVID-19 infections from chest X-ray images, Mahajan et al. (61) developed a single-shot MultiBox detector (SSD) in conjunction with deep transfer learning models and achieved high precision (i.e., 93.01%).

The authors of (60) used transfer learning to diagnose COVID-19 over 1,326 chest X-ray images, and the final method, the 121-layer Dense Convolutional Network (COVID-CXNet), was developed using the well-known CheXNet model (62). They achieved 99.04% accuracy using the COVID-CXNet method.

In (59), the authors conducted in-depth convolutional neural network (CNN) fine-tuning experiments and showed that models pre-trained on larger out-of-domain datasets demonstrate enhanced performance. Also, higher-quality images include more clinical information when the hyperparameters are chosen properly, and using mixups during training enhanced the model's performance.

According to the related works, to evaluate the performance of our proposed compact convolutional Transformer method, we took into account almost all of the standard and most important evaluation metrics, including accuracy (99.22%), precision (98.88%), recall (99.58%), F1-score (99.23%), AUC-ROC (99.27%), FPR 1.13, FNR (0.42%), and TNR (98.87%), which is outstanding in this regard. The results of our study show that this research is superior to many similar and state-of-the-art works in general and also when each of the evaluation metrics is considered or is completely comparable with them, and Table 10 confirms this claim.

6. Conclusion and future works

In this paper, a transformer-based model is proposed for screening chest X-ray images to detect COVID-19 disease. The proposed model is based on Compact Convolutional Transformers, whose main advantage over the other transformer-based models is its less need for data. This is important since in most medical domains data scarcity is ubiquitous. Using different metrics, we have demonstrated the efficacy of the proposed model for COVID-19 diagnosis. In future work, we tend to evaluate our proposed approach to other diseases related to human beings' lungs. That is to say, instead of classifying in a binary fashion positive and negative COVID-19, the approach should detect more classes of lung disorders.

Data availability statement

The dataset presented in this study can be found at https://www.kaggle.com/datasets/andyczhao/covidx-cxr2?select=competition_test.

Author contributions

JHJ, AM, and RL designed the study. AM performed the implementation of the approach. MM performed the literature review. AM and MAN wrote the methodology. JHJ, MAN, and RL did the discussion. JHJ, AM, and MAN edited the final version of the article. MAN supervised the project. RL co-supervised the study. All authors have read and approved the final manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Elibol E. Otolaryngological symptoms in COVID-19. *Eur Arch Otorhinolaryngol.* (2021) 278:1233–6. doi: 10.1007/s00405-020-06319-7

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

3. Chen X, Laurent S, Onur OA, Kleineberg NN, Fink GR, Schweitzer F, et al. A systematic review of neurological symptoms and complications of COVID-19. *J Neurol.* (2021) 268:392–402. doi: 10.1007/s00415-020-10067-3

4. Jin Y, Yang H, Ji W, Wu W, Chen S, Zhang W, et al. Virology, epidemiology, pathogenesis, and contr of COVID-19. *Viruses.* (2020) 12:372. doi: 10.3390/v12040372

5. Ismael AM, Sengür A. Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Syst Appl.* (2021) 164:114054. doi: 10.1016/j.eswa.2020.114054

6. Chai J, Zeng H, Li A, Ngai EW. Deep learning in computer vision: a critical review of emerging techniques and application scenarios. *Mach Learn Appl.* (2021) 6:100134. doi: 10.1016/j.mlwa.2021.100134

7. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: *2017 International Conference on Engineering Technology (ICET)* (2017). IEEE. p. 1–6.

8. O'Shea K, Nash R. An introduction to convolutional neural networks. *arXiv preprint.* (2015).

9. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Commun ACM.* (2017) 60:84–90. doi: 10.1145/3065386

10. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging.* (2018) 9:611–29. doi: 10.1007/s13244-018-0639-9

11. Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. *Neurocomputing.* (2017) 234:11–26. doi: 10.1016/j.neucom.2016.12.038

12. Masi I, Wu Y, Hassner T, Natarajan P. Deep face recognition: A survey. In: *2018 31st SIBGRAPI conference on graphics, patterns images (SIBGRAPI)* (2018). IEEE. p. 471–8.

13. Kortli Y, Jridi M, Al Falou A, Atri M. Face recognition systems: a survey. *Sensors.* (2020) 20:342. doi: 10.3390/s20020342

14. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. *BMC Med Imaging.* (2022) 22:1–13. doi: 10.1186/s12880-022-00793-7

15. Garcia-Garcia A, Orts-Escolano S, Oprea S, Villena-Martinez V, Martinez-Gonzalez P, Garcia-Rodriguez J. A survey on deep learning techniques for image and video semantic segmentation. *Appl Soft Comput.* (2018) 70:41–65. doi: 10.1016/j.asoc.2018.05.018

16. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-art natural language processing. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics (2020). p. 38–45.

17. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Huggingface's transformers: state-of-the-art natural language processing. *arXiv preprint.* (2019). doi: 10.18653/v1/2020.emnlp-demos.6

18. Edunov S, Ott M, Auli M, Grangier D. Understanding back-translation at scale. *arXiv preprint.* (2018). doi: 10.18653/v1/D18-1045

19. Abbasiantaeb Z, Momtazi S. Text-based question answering from information retrieval and deep neural network perspectives: a survey. *Wiley Interdiscip Rev Data Mining Knowledge Discov.* (2021) 11:e1412. doi: 10.1002/widm.1412

20. Habimana O, Li Y, Li R, Gu X, Yu G. Sentiment analysis using deep learning approaches: an overview. *Sci China Inform Sci.* (2020) 63:1–36. doi: 10.1007/s11432-018-9941-6

21. Mathew L, Bindu V. A review of natural language processing techniques for sentiment analysis using pre-trained models. In: *2020 Fourth International Conference on Computing Methodologies Communication (ICCMC)* (2020). IEEE. p. 340–5.

22. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. *ACM Computing Surveys.* (2021) 54:1–41. doi: 10.1145/3505244

23. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv Preprint.* (2020).

24. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal.* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005

25. 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) 10:1–12. doi: 10.1038/s41598-020-76550-z

26. Marques G, Agarwal D, de la Torre Díez I. Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. *Appl Soft Comput.* (2020) 96:106691. doi: 10.1016/j.asoc.2020.106691

27. Singh M, Bansal S, Ahuja S, Dubey RK, Panigrahi BK, Dey N. Transfer learning–based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data. *Med Biol Eng Comput.* (2021) 59:825–39. doi: 10.1007/s11517-020-02299-2

28. Islam MZ, Islam MM, Asraf A. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inform Med Unlocked.* (2020) 20:100412. doi: 10.1016/j.imu.2020.100412

29. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Anal Appl.* (2021) 24:1207–20. doi: 10.1007/s10044-021-00984-y

30. Goel T, Murugan R, Mirjalili S, Chakrabartty DK. OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. *Appl Intell.* (2021) 51:1351–66. doi: 10.1007/s10489-020-01904-z

31. Al Rahhal MM, Bazi Y, Jomaa RM, AlShibli A, Alajlan N, Mekhalfi ML, et al. Covid-19 detection in ct/x-ray imagery using vision transformers. *J Pers Med.* (2022) 12:310. doi: 10.3390/jpm12020310

32. Mondal AK, Bhattacharjee A, Singla P, Prathosh A. xViTCOS: explainable vision transformer based COVID-19 screening using radiography. *IEEE J Transl Eng Health Med.* (2021) 10:1–10. doi: 10.1109/JTEHM.2021.3134096

33. Liu C, Yin Q. Automatic diagnosis of covid-19 using a tailored transformer-like network. In: *Journal of Physics: Conference Series*, Vol. 2010 (2021). IOP Publishing. p. 012175.

34. Zhang L, Wen Y. A transformer-based framework for automatic COVID19 diagnosis in chest CTs. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021). p. 513–8.

35. Konar D, Panigrahi BK, Bhattacharyya S, Dey N, Jiang R. Auto-diagnosis of COVID-19 using lung CT images with semi-supervised shallow learning network. *IEEE Access.* (2021) 9:28716–28. doi: 10.1109/ACCESS.2021.3058854

36. Vaid S, Kalantar R, Bhandari M. Deep learning COVID-19 detection bias: accuracy through artificial intelligence. *Int Orthop.* (2020) 44:1539–42. doi: 10.1007/s00264-020-04609-7

37. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Comput Biol Med.* (2020) 121:103792. doi: 10.1016/j.compbiomed.2020.103792

38. Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos Solitons Fractals.* (2020) 138:109944. doi: 10.1016/j.chaos.2020.109944

39. Ahuja S, Panigrahi BK, Dey N, Rajinikanth V, Gandhi TK. Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices. *Appl Intell.* (2021) 51:571–85. doi: 10.1007/s10489-020-01826-w

40. Sharifrazi D, Alizadehsani R, Roshanzamir M, Joloudari JH, Shoeibi A, Jafari M, et al. Fusion of convolution neural network, support vector machine and Sobel filter for accurate detection of COVID-19 patients using X-ray images. *Biomed Signal Process Control.* (2021) 68:102622. doi: 10.1016/j.bspc.2021.102622

41. Khozeimeh F, Sharifrazi D, Izadi NH, Joloudari JH, Shoeibi A, Alizadehsani R, et al. Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. *Sci Rep.* (2021) 11:1–18. doi: 10.1038/s41598-021-93543-8

42. Kumar A, Tripathi AR, Satapathy SC, Zhang Y-D. SARS-Net: COVID-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network. *Pattern Recognit.* (2022) 122:108255. doi: 10.1016/j.patcog.2021.108255

43. Esmi N, Golshan Y, Asadi S, Shahbahrami A, Gaydadjiev G. A fuzzy fine-tuned model for COVID-19 diagnosis. *Comput Biol Med.* (2023) 2023:106483. doi: 10.1016/j.compbiomed.2022.106483

44. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst*. (2017) 30:1–11.

45. Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv preprint*. (2018). doi: 10.5555/3295222.3295349

46. Hassannataj Joloudari J, Azizi F, Nematollahi MA, Alizadehsani R, Hassannatajjeloudari E, Nodehi I, et al. GSVMA: a genetic support vector machine ANOVA method for CAD diagnosis. *Front Cardiovasc Med.* (2022) 8:2178. doi: 10.3389/fcvm.2021.760178

47. Wang X, Deng X, Fu Q, Zhou Q, Feng J, Ma H, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans Med Imaging.* (2020) 39:2615–2625. doi: 10.1109/TMI.2020.2995965

48. Prokop M, Van Everdingen W, van Rees Vellinga T, Quarles van Ufford H, Stöger L, Beenen L, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation. *Radiology.* (2020) 296:E97–104. doi: 10.1148/radiol.2020201473

49. Cai H, Zhu L, Han S. Proxylessnas: direct neural architecture search on target task and hardware. *arXiv preprint*. (2018).

50. Huang XS, Perez F, Ba J, Volkovs M. Improving transformer optimization through better initialization. In: *International Conference on Machine Learning* (2020). PMLR. p. 4475–83.

51. Alakus TB, Turkoglu I. Comparison of deep learning approaches to predict COVID-19 infection. *Chaos Solitons Fractals.* (2020) 140:110120. doi: 10.1016/j.chaos.2020.110120

52. Oguz Ç, Yaganoglu M. Detection of COVID-19 using deep learning techniques and classification methods. *Inf Process Manag.* (2022) 59:103025. doi: 10.1016/j.ipm.2022.103025

53. Srivastava G, Chauhan A, Jangid M, Chaurasia S. CoviXNet: A novel and efficient deep learning model for detection of COVID-19 using chest X-Ray images. *Biomed Signal Process Control.* (2022) 2022:103848. doi: 10.1016/j.bspc.2022.103848

54. Chen JI-Z. Design of accurate classification of COVID-19 disease in X-ray images using deep learning approach. *J ISMAC.* (2021) 3:132–48. doi: 10.36548/jismac.2021.2.006

55. Nasiri H, Hasani S. Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost. *Radiography.* (2022) 28:732–8. doi: 10.1016/j.radi.2022.03.011

56. Jain G, Mittal D, Thakur D, Mittal MK. A deep learning approach to detect Covid-19 coronavirus with X-ray images. *Biocybernetics Biomed Eng.* (2020) 40:1391–405. doi: 10.1016/j.bbe.2020.08.008

57. Al-Waisy A, Mohammed MA, Al-Fahdawi S, Maashi M, Garcia-Zapirain B, Abdulkareem KH, et al. COVID-DeepNet: hybrid multimodal deep learning system for improving COVID-19 pneumonia detection in chest X-ray images. *Comput Mater Continua.* (2021) 67:2409–29. doi: 10.32604/cmc.2021.012955

58. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Comput Biol Med.* (2020) 121:103795. doi: 10.1016/j.compbiomed.2020.103795

59. Zhao W, Jiang W, Qiu X. Fine-tuning convolutional neural networks for COVID-19 detection from chest X-ray images. *Diagnostics.* (2021) 11:1887. doi: 10.3390/diagnostics11101887

60. Haghanifar A, Majdabadi MM, Choi Y, Deivalakshmi S, Ko S. Covid-cxnet: detecting covid-19 in frontal chest x-ray images using deep learning. *Multimed Tools Appl.* (2022) 81:30615–45. doi: 10.1007/s11042-022-12156-z

61. Mahajan S, Raina A, Gao XZ, Pandit AK. COVID-19 detection using hybrid deep learning model in chest x-rays images. *Concurr Comput Pract Exp.* (2022) 34:e6747. doi: 10.1002/cpe.6747

62. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint*. (2017).