

Article

Detecting COVID-19 Effectively with Transformers and CNN-Based Deep Learning Mechanisms

Afamefuna Promise Umejiaku, Prastab Dhakal and Victor S. Sheng * 

Computer Science Department, Texas Tech University, Lubbock, TX 79409, USA;
afamefuna.p.umejiaku@ttu.edu (A.P.U.); prasdhak@ttu.edu (P.D.)

* Correspondence: victor.sheng@ttu.edu

Abstract: The COVID-19 pandemic has been a major global concern in the field of respiratory diseases, with healthcare institutions and partners investing significant resources to improve the detection and severity assessment of the virus. In an effort to further enhance the detection of COVID-19, researchers have investigated the performance of current detection methodologies and proposed new approaches that leverage deep learning techniques. In this article, the authors propose a two-step transformer model for the multi-class classification of COVID-19 images in a patient-aware manner. This model is implemented using transfer learning, which allows for the efficient use of pre-trained models to accelerate the training of the proposed model. The authors compare the performance of their proposed model to other CNN models commonly used in the detection of COVID-19. The experimental results of the study show that CNN-based deep learning networks obtained an accuracy in the range of 0.76–0.92. However, the proposed two-step transformer model implemented with transfer learning achieved a significantly higher accuracy of 0.9735 ± 0.0051 . This result indicates that the proposed model is a promising approach to improving the detection of COVID-19. Overall, the findings of this study highlight the potential of deep learning techniques, particularly the use of transfer learning and transformer models, to enhance the detection of COVID-19. These approaches can help healthcare institutions and partners to reduce the time and difficulty in detecting the virus, ultimately leading to more effective and timely treatment for patients.

Keywords: COVID-19; CNN; transformers; transfer learning



Citation: Umejiaku, A.P.; Dhakal, P.; Sheng, V.S. Detecting COVID-19 Effectively with Transformers and CNN-Based Deep Learning Mechanisms. *Appl. Sci.* **2023**, *13*, 4050. <https://doi.org/10.3390/app13064050>

Academic Editor: Jan Egger

Received: 18 February 2023

Revised: 16 March 2023

Accepted: 17 March 2023

Published: 22 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

COVID-19 is a contagious respiratory disease caused by the SARS-CoV-2 virus. It can cause pneumonia, which is an infection that inflames the air sacs in the lungs. COVID-19 pneumonia can be severe and life-threatening, particularly for older individuals or those with underlying health conditions. Symptoms of COVID-19 pneumonia include cough, fever, shortness of breath, chest pain, and fatigue. Early detection of COVID-19 is crucial, as the virus can progress rapidly, leading to the need for intensive care and respiratory support. Medical professionals have difficulty distinguishing COVID-19 pneumonia from other types of pneumonia based on physical symptoms alone, so accurate testing methods are essential for prompt diagnosis and treatment [1].

Following the outbreak of COVID-19, the world has faced significant challenges, and insufficient testing has been identified as one of the underlying causes of the rapid spread of the virus. One of the significant problems that medical professionals faced was the time it took to obtain results from the nucleic acid amplification tests (NAATs) and antigen tests. Although these tests are effective, with some versions, such as RT-PCR and RDT, they have proven insufficient due to the time it takes to get results and insufficient quantities of available tests [2]. To address this issue, researchers began exploring alternative testing methods to boost the detection of COVID-19. Computerized tomography (CT) scans are one of the alternative methods that have been utilized to aid in the rapid detection of COVID-19.

Medical professionals have used CT scans as a complementary tool for COVID-19 diagnosis to detect lung damage or other abnormalities that may suggest the presence of the virus. The use of CT scans for COVID-19 detection is a non-invasive and rapid method that can provide results in a matter of minutes.

Recent studies have shown that the severity of lung abnormalities quantified on chest CT can be used in the detection of COVID-19 and the evaluation of disease severity and progression. While concerns have been raised about the number of false positives and negatives associated with CT scans, it has become a preferred method of preliminary investigation for some medical facilities to detect COVID-19 even at early stages [3–5]. Saeed et al. [6] concluded that the 25-point CT severity score correlates well with the clinical severity of COVID-19, indicating that the chest CT scoring system can aid in predicting the presence of the COVID-19 disease. Mogami et al. [7] and Zhang et al. [8] proved that the patterns of CT findings, that is, the severity of lung abnormalities quantified on chest CT, can be used in the detection of COVID-19 and evaluation of the disease severity and progressive stages. Certain CT characteristics such as consolidation, GGO (ground-glass opacity), and crazy-paving signs have been identified as indicators for detecting and evaluating COVID-19 [9]. In general, although conventional viral testing approaches continue to be considered the most reliable method for diagnosing COVID-19, the use of CT scans has emerged as an important tool for the prompt detection and evaluation of the disease. However, interpreting these scans necessitates a high level of expertise in radiology, and there are limited numbers of professionals who possess this skill, which may lead to errors in the interpretation of results [4]. As a result, researchers have started exploring alternative, quicker, and more straightforward approaches for identifying COVID-19. One such approach involves utilizing machine learning techniques on both X-ray images and CT scans. Machine learning, which is a subset of artificial intelligence, automates the development of analytical models by examining data [10–12]. The subdivision of machine learning that creates models specifically for identifying COVID-19 through CT scans is commonly referred to as computer vision. In computer vision, one of the techniques that are frequently used is deep learning. It is a neural network architecture that incorporates a vast amount of labeled data to establish a model. Researchers have explored several neural network architectures in the context of COVID-19 detection, including ResNet50, Covid-ResNet, VGG16, VGG19, and DenseNet. ResNet50 and its more recent and refined versions, such as ResNet-rs and ResNet_V2, have shown significant success in identifying COVID-19 from CT scans. EfficientNet, a highly successful convolutional neural network architecture, has become an attractive option due to its ability to balance accuracy, speed, and computational efficiency. VGG16 and VGG19 are popular CNN architectures, while DenseNet utilizes a feed-forward model to address the issue of vanishing gradients by recycling features and reducing the number of parameters to be trained. Transformers, which were primarily used for natural language processing tasks, have seen increased use in computer vision and have also been used in COVID-19 detection [13–15].

To advance the application of artificial intelligence in viral detection, we took the following steps:

1. We conducted an extensive analysis of the available research on using CT scans for COVID-19 detection.
2. We explored different techniques for creating models that can detect COVID-19.
3. We developed an enhanced model for detecting the virus.

2. Materials and Methods

2.1. Dataset

In this study, we used a large COVID-19 CT scan dataset (Maftouni, 2021) [16], which is considered one of the most comprehensive and widely used datasets in the literature on COVID-19 diagnosis. The choice of dataset in any machine learning project is crucial as it has a direct impact on the performance of the models trained. The dataset was obtained from 7 public datasets available on Kaggle, which we carefully curated and preprocessed

to ensure that they contained only high-quality lung images that could provide valuable information about the state of the lung. To further enhance the quality of the dataset, we removed lung images that did not contain any information about the inside of the lung. This was done to ensure that the dataset was focused on images that could help in the detection and diagnosis of COVID-19. After preprocessing, the dataset consisted of a total of 17,104 images, which were divided into three categories: 7593 images from 466 patients with COVID-19, 6893 images from 604 patients without any signs of pneumonia, and 2618 images from 60 patients with community-acquired pneumonia (CAP). This large and diverse dataset provided a rich and representative sample of lung images, which could help to train and evaluate machine learning models for COVID-19 diagnosis.

Figure 1 in the study provides a visual representation of the different categories of images present in the dataset. It shows a sample of a COVID-19-infected CT scan, a healthy CT scan, and a CT scan of a lung infected with community-acquired pneumonia (CAP), highlighting the differences in the appearance of these scans. The availability of such a large and diverse dataset allowed us to train and evaluate machine learning models with high accuracy and confidence, providing a valuable tool in the fight against COVID-19.

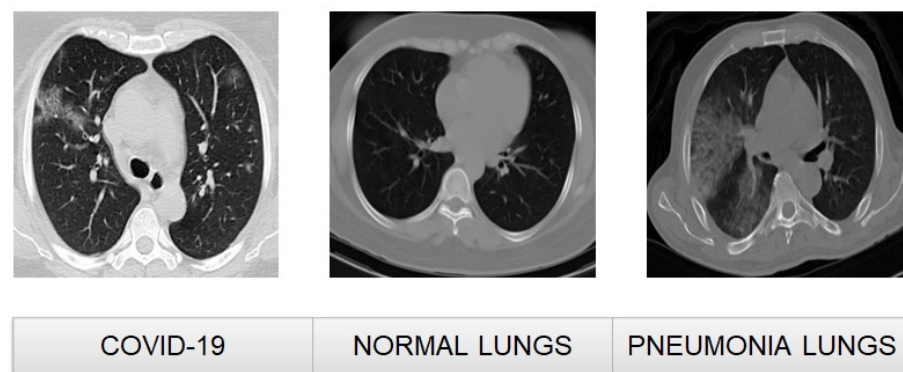


Figure 1. Examples of COVID-19 Infected, Normal, and Pneumonia Lung Images in the Large COVID-19 CT Scan Dataset.

2.2. Dataset Processing and Splitting

To prepare the dataset for machine learning, we transformed the data into a format that can be easily used for model training. The images in our dataset underwent the following transformations:

1. The first step in this process was to resize the images to a standard size and convert them to grayscale. This helps to simplify the processing steps that followed and ensured that all images were of the same size and color format, as demonstrated in Figure 2A.
2. Next, binary thresholding was applied to the images, which involved converting them to black and white and highlighting key areas of the images of importance for the contouring algorithm. This step simplified the detection and isolation of relevant features in the images.
3. The OpenCV library's findContours function was then used to detect all the contours present in the images, as shown in Figure 2B. This allowed for the identification of the different shapes and patterns present in the images, which is an important step in many image processing tasks.
4. The boundary contour was determined using the outer lung region as a reference, as illustrated in Figure 2C. This contour was used to crop the original images, which were already resized to a standard size of 256×256 pixels, as can be seen in Figure 2D with a white background and in Figure 2E with a black background.
5. Finally, the images were normalized and converted to a uniform depth of 8-bit, which ensured that the pixel values were consistent across all images. This step is

important for ensuring that the data are well-prepared for the training of machine learning models, which often require data to be in a specific format.

Algorithm 1 outlines the complete procedure of dataset preparation, with Figure 2 showing the images as they were captured and cropped using the contours detected. By following these steps, the dataset was transformed into a format that was easily usable for training machine learning models, allowing for more accurate and robust results.

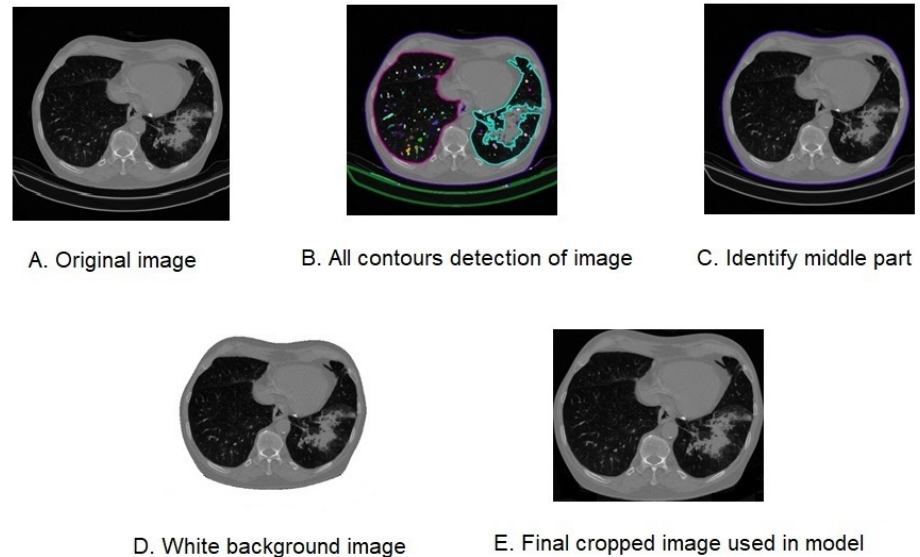


Figure 2. Images Illustrating Various Processing Steps and Outputs.

Algorithm 1: Data preparation Algorithm

Input: Image dataset of 512×512 from Kaggle collected by Maftouni, et al 2021 [16]

Output: 256×256 cropped images split in patient aware manner

while *not final image* **do**

 read images to numpy array;

 get binary threshold of image;

 get the contours;

 get boundary contours;

 draw the boundary

if *contour with max area* **then then**

 select the contour;

 return contour;

else

 keep comparing contours to find the max contour;

end

 crop images based on contour;

end

The process of splitting the dataset for model training and testing was carried out in a patient-aware manner, meaning that no individual patient's data were separated into different test and training folders. This approach helps to ensure that the model is evaluated on data from patients that it has not seen during training. Following the methodology of Maftouni et al. [16], we separated the dataset into a test and training dataset. The training dataset consisted of 15,871 images, with 7064 images showing COVID-19, 6387 images with no sign of pneumonia, and 2420 images with community-acquired pneumonia. The training samples were further divided into training and validation samples using a 10% test–train split for the evaluation process. This process involved randomly dividing the data into a training set and a validation set, with the model trained on the former and evaluated on the latter. The test dataset consisted of 1233 images, with 529 images showing

COVID-19, 506 images with no sign of pneumonia, and 198 images with community-acquired pneumonia. To increase the size of the dataset and help the model to generalize better, we performed data augmentation on the images. This involved applying a range of data transforms, such as center cropping, horizontal flipping, and random resizing. These transforms created new images that were similar to the original images but with slight variations in position, orientation, and size. By using these augmented images during training, we can increase the size of the dataset and expose the model to a wider range of image variations, helping it to better generalize to new, unseen images. Through this process of data splitting and augmentation, we were able to create the curated CT COVID-19 dataset that we used for training and testing our machine learning models.

2.3. Existing Model Recreation

We approached the task of developing our proposed machine learning models for COVID-19 detection using our curated dataset. First, we utilized several existing model architectures that have been proven effective for image classification tasks. The models used were DenseNet-121 and VGG-19, both of which have been utilized in various classification tasks, including COVID-19 detection. Additionally, we recreated the Maftouni et al. [16] ensemble model, which is known to achieve higher accuracy compared to a single model. To do this, we trained the curated dataset using two versions of the Maftouni Ensemble model: one with fully connected (FC) layers and another with an FC+ support-vector machine (SVM). For the VGG-19 and DenseNet-121 models, we implemented average pooling and pre-trained ImageNet weights using functions provided in the TensorFlow library. The use of pre-trained weights allows for the transfer of learned features from a large dataset of labeled images such as ImageNet to the new task of COVID-19 detection. Leveraging pre-trained weights, models can learn to recognize COVID-19-related features from a smaller, curated dataset, which improves its performance. By implementing average pooling and pre-trained ImageNet weights, we were able to improve the model's performance, which ultimately increased the accuracy of COVID-19 detection.

2.4. Vision Transformer with a Custom Linear Classifier

To carry out our multi-class classification task, we utilized the Vision Transformer from the Hugging Face library, a powerful tool in natural language processing and computer vision. The creators of this tool customized a linear classifier with a dense layer, which was designed to classify the output of the transformer encoder block into the specific classes that they were interested in. To prepare their dataset for the ViT model, they employed the ViTFeatureExtractor, which allowed us to resize and normalize their curated dataset in a way that was suitable for the ViT model.

The ViT model was pre-trained on the massive imagenet21k dataset, which contains over a million images and thousands of classes. To adapt the ViT model to their particular classification task, we converted the input images into flattened patches using a patch size of 16. This allowed the model to learn from local features within each image while also retaining information about the larger context. To fine-tune the Vision Transformer, we used both one-step and two-step transfer learning approaches. We set the training arguments to a learning rate of 0.00005 with 50 epochs and an early stopping patience of 10. To optimize the training process, we experimented with different parameters, including device training batch sizes, device evaluation batch sizes, weight decay, gradient accumulation steps, and the warmup ratio. To validate the results, we ran the same training on five different folds and calculated the mean and standard deviation of the results. We found that both the one-step and two-step transfer learning approaches produced different results, which we discuss in greater detail in Section 3. Overall, the Vision Transformer proved to be a valuable tool for the multi-class classification task, offering high accuracy and a strong performance on a range of metrics.

2.4.1. One-Step Transfer Learning Approach

In the one-step approach, we utilized our pre-trained Vision Transformer model along with our customized linear classifier to fine-tune the model using our curated dataset, as depicted in Figure 3. The aim was to classify the input images into three classes, corresponding to our fine-tuning dataset. The fine-tuning process involved adjusting the model's pre-existing parameters to better suit our target task while keeping its fundamental structure intact. This enabled the model to learn new features that were specifically tailored to our classification task. To assess the performance of the model, we employed five-fold cross-validation and computed the mean and standard deviation. The details of the results obtained using this approach are discussed in Section 3.

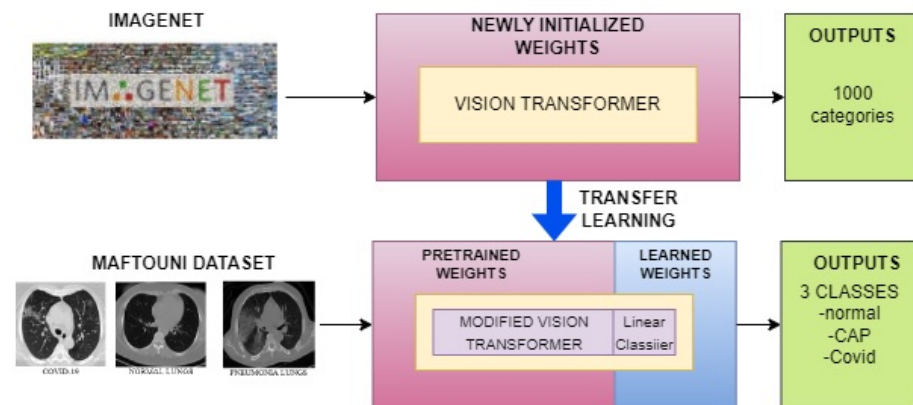


Figure 3. One-step Transfer Learning approach.

2.4.2. Two-Step Transfer Learning Approach

In the two-step transfer learning approach, we first fine-tuned the pre-trained Vision Transformer model on a chest X-ray image (pneumonia) dataset (PAUL MOONEY), which consists of X-ray images labeled as normal or pneumonia. This allowed us to incorporate additional relevant information into our pre-trained model, making it more suitable for our multi-class classification task. We then customized the pre-trained model with our own linear classifier and further fine-tuned the model using our curated dataset, which includes COVID-19, pneumonia, and normal chest X-ray images. The fine-tuning process using the curated dataset allowed us to improve the model's performance on our specific classification task. By fine-tuning the pre-trained model with our dataset, we were able to leverage the knowledge already present in the pre-trained model and adapt it to our specific use case. This allowed us to achieve better results than if we had started training the model from scratch. The use of two-step transfer learning has been shown to be effective in various computer vision tasks, particularly in cases where the target dataset is small and may not have enough data to train a model from scratch. In our case, the COVID-19 dataset was limited, the two-step transfer learning approach allowed us to overcome this challenge and achieve better results. Figure 4 illustrates the process we followed in the two-step transfer learning approach, highlighting the steps involved in fine-tuning the pre-trained model on the pneumonia dataset and then further customizing and fine-tuning it using our curated dataset.

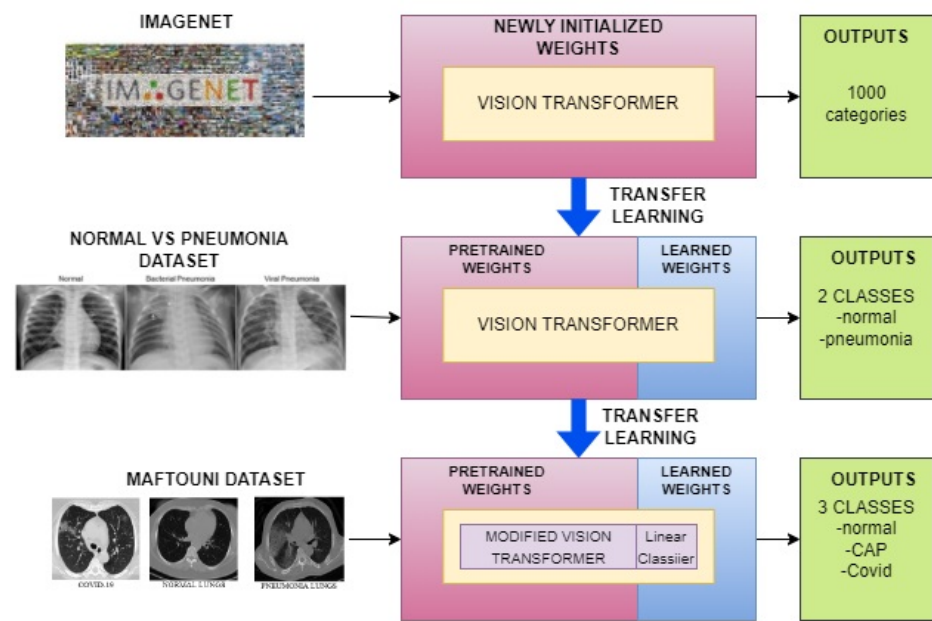


Figure 4. Two-step Transfer Learning approach.

3. Results

During the fine-tuning process, we performed extensive experiments to determine the optimal hyperparameters for our Vision Transformer models. These hyperparameters included training batch size, evaluation batch size, learning rate, weight decay, gradient accumulation steps, and warmup ratio. To achieve the best results, we systematically varied these hyperparameters and evaluated the performance of the model using the validation set. Through this process, we found that the best hyperparameters for our models were a training batch size of 8, evaluation batch size of 4, weight decay of 0.01, gradient accumulation step of 4, and warmup ratio of 0.1. Using these optimal hyperparameters, we evaluated the performance of our models on our curated dataset. We trained each model using a 5-fold cross-validation approach, and the results of our evaluation are shown in Table 1.

Table 1. Representation of average model performance with standard deviation

Model	Precision	Recall	F1-Score	Mean(Acc)
VGG-19	0.8001 ± 0.0097	0.7684 ± 0.0094	0.7439 ± 0.0188	0.7660 ± 0.0111
DenseNet-121	0.8777 ± 0.0071	0.9409 ± 0.0098	0.9093 ± 0.0110	0.9228 ± 0.0035
Maftouni Ensemble Model with FC	0.8984 ± 0.0071	0.9830 ± 0.0071	0.9386 ± 0.0082	0.9508 ± 0.0012
Maftouni Ensemble Model with FC+SVM	0.9078 ± 0.0110	0.9789 ± 0.0103	0.9420 ± 0.0097	0.9529 ± 0.0043
Vision Transformer with 1 step Classifier Model	0.9698 ± 0.0054	0.9666 ± 0.0075	0.9618 ± 0.0119	0.9693 ± 0.0037
Vision Transformer with 2-step Classifier Model	0.9766 ± 0.0081	0.9761 ± 0.0037	0.9767 ± 0.0075	0.9736 ± 0.0051

The VGG-19 model achieved an F1-score of 0.7439 and an average accuracy of 0.7660, while the DenseNet-121 model had an F1-score of 0.9093 and an average accuracy of 0.9228. The Maftouni Ensemble model with FC had an F1-score of 0.9386 and an average accuracy of 0.9508, and the Maftouni Ensemble model with FC+ SVM had an F1-score of 0.9420 and an average accuracy of 0.9529. The Vision Transformer model with a one-step classifier had an F1-score of 0.9618 and an average accuracy of 0.9693, and the Vision Transformer model with a two-step classifier achieved the highest performance, with an F1-score of 0.9767 and an average accuracy of 0.9736. Overall, the results of our experiments demonstrate that the

Vision Transformer model with a two-step classifier is the most effective approach for the multi-class classification of COVID-19 chest X-ray images. However, we also showed that other models, such as DenseNet-121 and Maftouni Ensemble models, can also achieve high performance with our curated dataset.

4. Discussion

Farooq and Hafeez's 2020 study [17] employed fine-tuning techniques on a modified version of ResNet50 architecture, resulting in the creation of Covid-ResNet, a model designed to identify COVID-19 from CT scans. Similarly, in 2021, Sreejith and George [18] utilized ResNet50 architecture with a chest X-ray dataset to construct a machine learning model capable of detecting COVID-19. Researchers have also investigated the use of EfficientNet, a highly successful convolutional neural network architecture, either independently or as part of an ensemble network, due to its flexibility and versatility in customization to meet specific research and practitioner needs [19]. Other researchers have utilized varying forms of VGG16 and VGG19 models, either independently or in an ensemble, to create automated COVID-19 detection systems [20,21]. DenseNet, another deep neural network that employs a feed-forward model to address vanishing gradients, has also demonstrated exceptional performance in various studies [22,23]. Using Chouat et al. [24] comprehensive analysis of a range of popular models suggested that the VGG-19 model achieved the highest accuracy and F1 score of 0.905. Additionally, we referenced Maftouni et al. [16], where the models achieved an F1 score of 0.9423 and an accuracy of 0.9531. We evaluated several models for COVID-19 image classification, including VGG-19, DenseNet-121, Maftouni Ensemble with FC, and Maftouni Ensemble with FC+ SVM. We found that the performance of all the trained models met or exceeded our expectations, demonstrating the effectiveness of our curated dataset for model training.

Based on the suggestion that CNN-based models may struggle with the long-range relationship problem resulting from regional locality in convolution operations [25], we developed a model using Vision Transformer, an image classification model inspired by NLP transformers, which has been pre-trained on large datasets to address this issue. Several studies have indicated that transformers show promise as an alternative to traditional CNN-based neural networks [26–30]. Our model produced impressive results, achieving an F1 score of over 0.96 for the one-step classifier and over 0.97 for the two-step classifier. These findings demonstrate the high accuracy of our developed model for COVID-19 image classification and support the use of artificial intelligence in CT scans for COVID-19 detection.

As a future task, we suggest implementing multitask learning to identify not only whether a person has COVID-19, but also other potential issues. Multitask learning is a promising approach that allows for the simultaneous detection of multiple diseases or medical conditions. By leveraging the vast amounts of medical data available, AI can help clinicians make more comprehensive and accurate diagnoses, ultimately improving patient outcomes. Our study provides important insights into the potential of AI in healthcare and demonstrates the effectiveness of our model design for COVID-19 image classification. The use of AI in medical diagnosis has the potential to revolutionize healthcare, and we believe that our findings could inspire further research and development in this field. With the ongoing COVID-19 pandemic and the increasing demand for accurate and efficient medical diagnosis, the development and implementation of AI in healthcare is more important than ever.

5. Conclusions

In conclusion, our proposed method for COVID-19 detection outperformed state-of-the-art ensemble techniques in accurately identifying COVID-19 patients. Our study produced models with reasonable performance results, which we hope will encourage the increased use of artificial intelligence in healthcare. The success of our two-step transfer learning approach in improving model accuracy is a promising development, and it would

be interesting to see if this approach can be applied to other image classification tasks. Using a Vision Transformer, we were able to predict COVID-19 pneumonia in patients with an accuracy of 0.9736 ± 0.0051 . Our findings demonstrate the potential of artificial intelligence to aid in the diagnosis and treatment of COVID-19, and we believe that our method could be implemented in clinical settings to assist medical professionals in the identification and management of the disease. Furthermore, our study highlights the importance of a high-quality dataset in producing accurate and reliable results. We observed improved performance when training prior models with our carefully curated dataset, emphasizing the need for thorough data curation in artificial intelligence research. Overall, our study provides valuable insights into the potential of artificial intelligence in healthcare, and we hope that our findings will inspire further research in this field.

Author Contributions: A.P.U., P.D. and V.S.S. conceived of the presented idea. P.D. and A.P.U. worked on the methodology, and P.D. performed the experiment. P.D. and V.S.S. verified the modeling technique, and formal analysis and investigation were done by A.P.U. and P.D. Data curation was performed by P.D., and writing, original draft preparation, review, and editing were done by A.P.U. Visualizations were created by P.D., under the supervision of V.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request due to privacy/ethical restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lotfi, M.; Hamblin, M.R.; Rezaei, N. COVID-19: Transmission, prevention, and potential therapeutic opportunities. *Clin. Chim. Acta* **2020**, *508*, 254–266. [CrossRef] [PubMed]
2. Rubino, S.; Kelvin, N.; Bermejo-Martin, J.F.; Kelvin, D. As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. *J. Infect. Dev. Ctries.* **2020**, *14*, 265–267. [CrossRef] [PubMed]
3. Kwee, T.C.; Kwee, R.M. Chest CT in COVID-19: What the radiologist needs to know. *RadioGraphics* **2020**, *40*, 1848–1865. [CrossRef] [PubMed]
4. Hani, C.; Trieu, N.H.; Saab, I.; Dangeard, S.; Bennani, S.; Chassagnon, G.; Revel, M.-P. COVID-19 pneumonia: A review of typical CT findings and differential diagnosis. *Diagn. Interv. Imaging* **2020**, *101*, 263–268. [CrossRef] [PubMed]
5. Parasher, A. COVID-19: Current Understanding of Its Pathophysiology, Clinical Presentation and Treatment. Available online: <https://pmj.bmj.com/content/97/1147/312.citation-tools> (accessed on 3 December 2022).
6. Saeed, G.A.; Gaba, W.; Shah, A.; Al Helali, A.A.; Raidullah, E.; Al Ali, A.B.; Elghazali, M.; Ahmed, D.Y.; Al Kaabi, S.G.; Almazrouei, S. Correlation between chest CT severity scores and the clinical parameters of adult patients with COVID-19 pneumonia. *Radiol. Res. Pract.* **2021**, *2021*, 6697677. [CrossRef]
7. Mogami, R.; Lopes, A.J.; Araújo Filho, R.C.; de Almeida, F.C.S.; Messeder, A.M.D.C.; Koifman, A.C.B.; Guimarães, A.B.; Monteiro, A. Chest Computed Tomography in COVID-19 Pneumonia: A Retrospective Study of 155 Patients at a University Hospital in Rio de Janeiro, Brazil. Available online: <https://pubmed.ncbi.nlm.nih.gov/33583973/> (accessed on 2 December 2022).
8. Zhang, B.; Zhang, J.; Chen, H.; Chen, L.; Chen, Q.; Li, M.; Chen, Z.; You, J.; Yang, K.; Zhang, S. Novel coronavirus disease 2019 (COVID-19): Relationship between chest CT scores and laboratory parameters. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *47*, 2083–2089. [CrossRef] [PubMed]
9. Liao, J.; Chen, Y.; Huang, C.Q.; He, G.; Du, J.C.; Chen, Q.L. Clinical differences in chest CT characteristics between the progression and remission stages of patients with Covid-19 pneumonia. *Int. J. Clin. Pract.* **2020**, *75*, e13760. [CrossRef] [PubMed]
10. Fusco, R.; Grassi, R.; Granata, V.; Setola, S.V.; Grassi, F.; Cozzi, D.; Pecori, B.; Izzo, F.; Petrillo, A. Artificial Intelligence and COVID-19 using chest CT scan and chest X-ray images: Machine learning and deep learning approaches for diagnosis and treatment. *J. Pers. Med.* **2021**, *11*, 993. [CrossRef] [PubMed]
11. Zhu, Z.; Xingming, Z.; Tao, G.; Dan, T.; Li, J.; Chen, X.; Li, Y.; Zhou, Z.; Zhang, X.; Zhou, J.; et al. Classification of COVID-19 by compressed chest CT image through deep learning on a large patients cohort. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*, 73–82. [CrossRef] [PubMed]
12. El Naqa, I.; Murphy, M.J. What Is Machine Learning? In *Machine Learning in Radiation Oncology*; El Naqa, I., Li, R., Murphy, M., Eds.; Springer: Cham, Switzerland, 2015. [CrossRef]

13. Gunraj, H.; Sabri, A.; Koff, D.; Wong, A. Covid-net CT-2: Enhanced Deep Neural Networks for detection of COVID-19 from chest CT images through bigger, more diverse learning. *Front. Med.* **2022**, *8*, 3126. [[CrossRef](#)] [[PubMed](#)]
14. Kogilavani, S.V.; Prabhu, J.; Sandhiya, R.; Kumar, M.S.; Subramaniam, U.S.; Karthick, A.; Muhibbullah, M.; Imam, S.B. COVID-19 detection based on lung CT scan using Deep Learning Techniques. *Comput. Math. Methods Med.* **2022**, *2022*, 7672196. [[CrossRef](#)] [[PubMed](#)]
15. Zhao, W.; Jiang, W.; Qiu, X. Deep learning for COVID-19 detection based on CT Images. *Sci. Rep.* **2021**, *11*, 14353. [[CrossRef](#)] [[PubMed](#)]
16. Maftouni, M.; Law, A.C.C.; Shen, B.; Kong, Z.; Zhou, Y.; Yazdi, N.A. A Robust Ensemble-Deep Learning Model for COVID-19 Diagnosis based on an Integrated CT Scan Images Database. In *IIE Annual Conference*; Institute of Industrial and Systems Engineers: Peachtree Corners, GA, USA, 2021; pp. 632–637.
17. Farooq, M.; Hafeez, A. Covid-ResNet: A Deep Learning Framework for Screening of COVID-19 from Radiographs. Available online: <https://arxiv.org/abs/2003.14395> (accessed on 2 December 2022).
18. Sreejith, V.; George, T. Detection of COVID-19 from chest X-rays using resnet-50. *J. Phys. Conf. Ser.* **2021**, *1937*, 012002. [[CrossRef](#)]
19. Diallo, P.A.; Ju, Y. Accurate detection of covid-19 using K-efficientnet deep learning image classifier and K-COVID chest X-Ray Images Dataset. In proceedings of the 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 11–14 December 2020;. [[CrossRef](#)]
20. Chowdhury, N.K.; Kabir, M.A.; Rahman, M.M.; Rezoana, N. ECOVNet: A highly effective ensemble based deep learning model for detecting COVID-19. *PeerJ Comput. Sci.* **2021**, *7*, e551. [[CrossRef](#)] [[PubMed](#)]
21. Mishra, N.K.; Singh, P.; Joshi, S.D. Automated detection of COVID-19 from CT scan using Convolutional Neural Network. *Biocybern. Biomed. Eng.* **2021**, *41*, 572–588. [[CrossRef](#)] [[PubMed](#)]
22. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. Available online: <https://arxiv.org/abs/1608.06993> (accessed on 1 December 2022).
23. Zhang, Y.-D.; Satapathy, S.C.; Zhang, X.; Wang, S.-H. COVID-19 diagnosis via DenseNet and optimization of transfer learning setting. *Cogn. Comput.* **2021**. [[CrossRef](#)]
24. Chouat, I.; Echtioui, A.; Khemakhem, R.; Zouch, W.; Ghorbel, M.; Hamida, A.B. COVID-19 detection in CT and CXR images using Deep Learning Models. *Biogerontology* **2022**, *23*, 65–84. [[CrossRef](#)] [[PubMed](#)]
25. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. Available online: <https://arxiv.org/abs/2102.04306v1> (accessed on 3 December 2022).
26. Sagar, A. Vitbis: Vision Transformer for Biomedical Image Segmentation. Available online: <https://arxiv.org/abs/2201.05920> (accessed on 3 December 2022).
27. Schlemper, J.; Oktay, O.; Schaap, M.; Heinrich, M.; Kainz, B.; Glocker, B.; Rueckert, D. Attention Gated Networks: Learning to leverage salient regions in medical images. *Med. Image Anal.* **2019**, *53*, 197–207. [[CrossRef](#)] [[PubMed](#)]
28. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. Available online: <https://arxiv.org/abs/2010.11929v1> (accessed on 2 December 2022).
29. Boesch, G. Vision Transformers (ViT) in Image Recognition—2022 Guide. Available online: <https://viso.ai/deep-learning/vision-transformer-vit/> (accessed on 3 December 2022).
30. Lutins, E. Ensemble Methods in Machine Learning: What are They and Why Use Them? Available online: <https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f> (accessed on 2 December 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.