

COVID-19 Diagnosis from Chest X-Ray Images Using Convolutional Neural Networks and Effects of Data Poisoning

Karthika Menon(✉), V. Khushi Bohra, Lakshana Murugan, Kavya Jaganathan, and Chamundeswari Arumugam

Department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai 603110, India
{karthika18070,khushibohra18075,lakshana18082, kavya17074}@cse.ssn.edu.in, chamundeswaria@ssn.edu.in

Abstract. At the end of 2019, a new type of virus called SARS-CoV-2 began spreading resulting in a global pandemic. As of June 2021, almost 175 million people were affected worldwide. Symptom-wise, it is very difficult to diagnose if a person has Covid or just a viral infection. But, taking a close look at chest X-Rays is extremely helpful in the diagnostic process. The proposed methodology in this paper helps in classification of chest X-Ray images into 3 categories: 'Covid', 'Viral' and 'Normal'. The dataset was created by integrating 3 pre-existing evergrowing datasets and the ResNet-18 model was adopted to train it. The experimental results show that the classification of the chest X-Ray images was done with an accuracy of 0.9648. An adversarial machine learning approach was employed to poison the train data after which the classification accuracy dropped to 0.8711.

Keywords: COVID-19 · Pneumonia · Chest X-Rays · Machine learning · CNN · ResNet-18 · Image classification · Medical imaging · Data poisoning · Data security

1 Introduction

The COVID-19 pandemic is one of the greatest health crises this world has faced till date. As of June 2021, around 175 million people have been affected by COVID-19 and 3.86 million people have lost their lives. Most patients who are suspected to have COVID-19 have symptoms like cough, fever, fatigue, anorexia, shortness of breath, and sputum production. Chest x-ray is the first diagnostic technique that has a significant role in COVID diagnosis. With the unfortunate spread of COVID 19, rapid imaging analysis will allow for quick digital detection of cases and prevent spread, eliminating the time required to generate radiological reports.

Nowadays, deep learning concepts are developing at a very fast rate. Deep Learning is a combination of machine learning methods mainly focused on feature

© Springer Nature Switzerland AG 2021
O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12957, pp. 508–521, 2021.
https://doi.org/10.1007/978-3-030-87013-3_38

extraction and image classification. Deep learning uses the process of convolution where a deep convolutional neural network is used for feature extraction. The network uses multiple layers that process non-linear information. Each layer involves a transformation of the data into a higher and more abstract level. More complex information is learned as the network is traversed deeper. Higher layers contain the information that is used for segregation and classification.

The continuous growth of COVID chest x-ray dataset allows us to use deep learning models to train and learn to detect COVID from these x-rays. Further, more information can be learned with regards to severity depending on scores and features learned from the training. One of the major issues about COVID diagnosis is that the symptoms of pneumonia and COVID-19 are almost the same. Image analysis of chest x-ray images aid doctors and healthcare workers significantly in differentiating between a COVID infection and pneumonia.

The purpose of this research is to segregate the chest x-Rays present in the dataset using the pre-trained convolutional neural network, ResNet-18. Classification is done into 3 classes namely- 'Normal', 'Viral', 'COVID'. The results are extremely encouraging and demonstrate how effective deep learning and transfer learning applications are, with CNNs for detection of abnormal X-ray images from normal ones and identifying which chest x-rays actually show signs of COVID-19.

Data poisoning is a method to corrupt or weaken the machine learning system by attacking it. The attack was performed in the form of data manipulation where the pre-existing training data was changed. In doing this, the system classifies wrong data into desired classes. This is done because tampering with the training data impacts the model's ability to classify data into correct classes. Since the dataset used to conduct this research is public, it is prone to manipulation by various sources. Hence, data poisoning was done to demonstrate the vulnerability of this model in case data poisoning occurs. In this paper, data manipulation of the dataset was employed as part of a data poisoning approach which resulted in a significant drop in test accuracy.

2 Literature Survey

COVID-19 detection using x-ray images would help physicians accelerate the diagnosis to a considerable extent and hence it is an active field of research. The following are few existing works related to this area of study.

COVID-19 pneumonia detection in chest x-ray images using transfer learning of convolutional neural networks [1] by Manop Phankokkruad chose VGG16, Xception, and Inception-Resnet-V2 pre-trained models for transfer learning. The comparative results drawn from this study indicate that VGG16 produced the highest training accuracy of 97.19%. Pneumonia diagnosis using chest x-ray images and machine learning [2] developed by Sara Lee Kit Yee et al. have evaluated the performance of the three classifiers namely kNN,SVM and NN. The study concluded that the accuracy of SVM turned out to be the highest with 93.1%. Automated pneumonia diagnosis using a customized sequential convolutional neural network [3] by Raheel Siddiqi used an 18 layer customised CNN model to classify the dataset into two classes- 'Normal' and 'Pneumonia'. This

customised CNN model had an accuracy of 0.9375 which proved to be greater than the state-of-the-art model.

YU-net lung segment image preprocess methods used for common chest diseases prediction [4] proposed by Haoxiong Yu et al. have developed a novel method to improve the accuracy of CNN models using YU-net cleaned data. COVID-19 detection from chest x-ray images using deep learning and convolutional neural networks [5] developed by Antonios Makris et al. classified the images into 3 classes- 'Normal' ,'Covid' and 'Pneumonia'. The model applied Transfer Learning and evaluated 9 well known pre-trained CNN models out of which VGG16 obtained the highest overall accuracy of about 0.9588. COVID-19 chest radiography images analysis based on integration of image preprocess, guided grad-CAM,machine learning and risk management [6] by Tsung-Chieh Lin et al. proves that ensemble approach improves the accuracy compared to individual CNN models.

A neuro-heuristic approach for recognition of lung diseases from X-ray images [7] proposed by Qiao Ke et al. uses simplified image descriptors to detect degenerated lung tissues in x-ray images and predict respiratory diseases. Bio-inspired methods modeled for respiratory disease detection from medical images [8] by Marcin Woźniak et al. proposed a bio-inspired methodology search for areas in x-ray image where diseased tissues are likely to appear and hence detect respiratory diseases from that. Small Lung Nodules Detection Based on Fuzzy-Logic and Probabilistic Neural Network With Bioinspired Reinforcement Learning [9] by Giacomo Capizzi et al. presented an evaluation model based on a composition of fuzzy system combined with a neural network. The results show high performances with sensitivity and specificity reaching almost 0.95 and 0.90, respectively, with an accuracy of 0.9256.

3 Methodology

3.1 Proposed Model

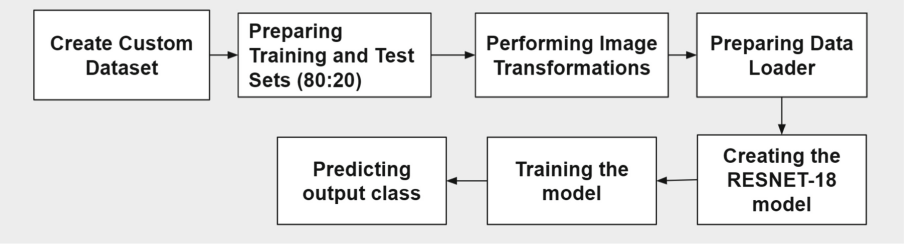


Fig. 1. Proposed model

The proposed model for solving the image classification problem in this paper is as detailed in Fig. 1. A custom dataset is first created by integrating 3 open source datasets (2 Kaggle and 1 GitHub) and then augmenting the images to obtain a balanced, more expansive database. Once the dataset is created, it is

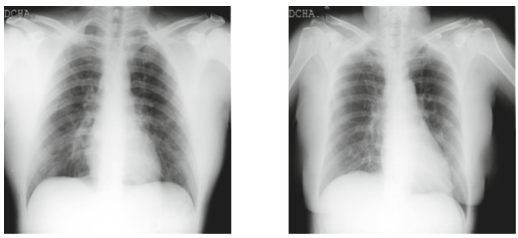
split into training and test sets applying an 80:20 train-to-test ratio. Python torchvision image transformations are then performed on the images to resize and customize them to desired dimensions before feeding them into the ResNet-18 model for training. The transformed images are imported into the data loader. The ResNet-18 model is then created and trained using the images in the train set by setting the hyperparameters. Finally, the model is tested using the images in the test set and the output classes are predicted i.e., Covid-19, Viral pneumonia or normal.

3.2 Data Description

The dataset used to build the model is a combination of 3 datasets. These include kaggle's COVID-19 radiography database [10], comprising 10,192 normal images, 3,616 COVID-19 images and 1,345 viral pneumonia images. The second dataset used is IEEE8023's covid chestxray dataset [11] from which 562 COVID-19 images, 40 viral pneumonia images and 18 normal images were incorporated into the dataset. The final dataset used is Kaggle's COVID-19 x-rays [12] dataset which consists of 71 COVID-19 images and 7 Normal images. In total they make up a dataset comprising 15,851 images i.e., 10,217 normal images, 4249 COVID-19 images and 1,385 viral pneumonia images (refer Fig. 2, 3, 4 for sample chest x-ray images).

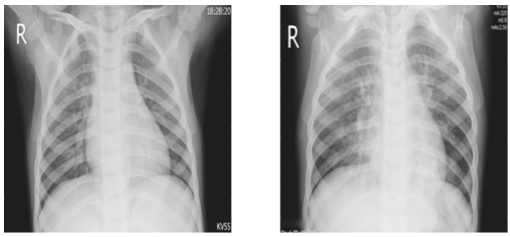
Several modules of the python library imgaug.augmenters including Fliplr (horizontal flip), Flipud (vertical flip), Gaussian Blur, Dropout, AdditiveGaussianNoise were applied to augment the dataset. The dataset after augmentation comprises 7940 covid-19 images, 8192 normal images and 7366 viral pneumonia images which make up a total of 23,498 images in the dataset. Python torchvision image transformation functions such as Resize, Normalize, ToTenser etc. were applied in order to customize the images to the necessary input specification (224×224) before feeding into the pretrained ResNet-18 model.

The data loader was prepared to obtain training and test datasets. The training dataset has a total of 23,488 images comprising 8192 normal images, 7366 viral examples and 7940 covid-19 samples. The test dataset has a total of 6000 images, consisting 2000 image samples from each of the 3 classes.



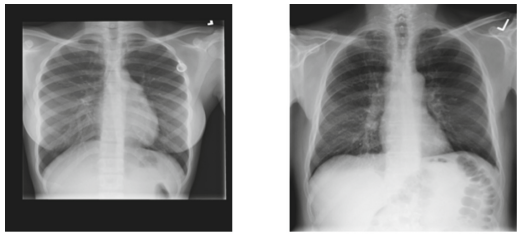
(a) **(b)**

Fig. 2. Examples of COVID-19 chest x-ray images



(a) (b)

Fig. 3. Examples of viral pneumonia chest x-ray images



(a) (b)

Fig. 4. Examples of normal chest x-ray images

3.3 Classification Method

A transfer learning approach of extracting features from a base convolutional neural network, and re-purposing feature maps learned previously on a larger dataset was employed to perform the classification task. The feature extraction task in this model was carried out using the convolutional layers of a pretrained ResNet-18 network.

The ResNet-18 architecture comprises alternating sequential convolutional layers, batch normalization layers and pooling layers. The convolutional layer generates feature maps based on the concentration of white regions in the chest x-ray indicating presence of infection. This is owing to the fact that the lungs of a covid-19 patient are full of water and sputum, hence more radiation is absorbed causing the affected region to appear in white or grey. Covid-19 and viral pneumonia chest x-rays differ in the distribution of lesions around the lungs. An axial distribution of lesions in the peripheral zone is observed in case of covid images, whereas a more random and diffuse distribution is observed in case of pneumonia x-rays.

The activation function applied to introduce non-linearity in the feature maps outputted by the convolutional layer is Rectified Linear Unit (RELU). In addition pooling layers are included to down sample the feature maps and prevent overfitting of the model, by making the model invariant to modulations in images brought forth by different lighting/positions. This improves the generalization of the model. Both max pooling and average pooling techniques were employed

to achieve the same. A total of 512 in-features were applied in the creation of the model. The final connected dense layers, output the predicted class based on the features extracted from the previous sequential layers, which in this case is a total of 3 out-features i,e. COVID-19, viral pneumonia and normal.

The code was executed on a Jupyter notebook. The steps involved in the implementation process were namely, (1) Creating a custom dataset by augmenting images, (2) Preparing the training and test sets, (3) Performing image transformations to customise model input (4) Preparing the data loader, (5) Data visualization, (6) Creating the model, (7) Training the model and finally (8) Classifying images and showing predictions.

3.4 Training

The decision on which architecture is most suitable to train the dataset was made based on repeated experimentation. A few of the prominent pre-trained network architectures for image classification tasks are as tabulated in Table 1.

Table 1. Pre-trained network architectures for image classification tasks

VGG16	AlexNet	ResNet-18
The input to this network is an image of dimensions 224*224*3	The network has an input image size of 227*227*3	The image input size is of 224*224
It has 16 convolution layers and consists of 138 million parameters	The total parameters in this architecture is 62.3 million	ResNet-18 is a convolution neural network that consists of 18 layers with around 11 million trainable parameters
It has 16 convolution layers and consists of 138 million parameters. It uses multiple layers that include kernels of 3*3 dimension for convolution and 2*2 size for max pool	The AlexNet architecture consists of 8 layers among which 5 are convolution layers with a combination of max pooling followed by 3 fully connected layers. ReLU activation is used in each layer except the output layer	It has convolution layers with filters of size 3*3 and 2 pooling layers

Among the 3, the ResNet-18 model proved to be easier to train owing to fewer trainable parameters as compared to the other pretrained networks. In addition, it gave optimum accuracy results at the end of training.

As part of the transfer learning approach of feature extraction, once the ResNet-18 model is created, the last fully connected (FC) layer of the pre-trained model is modified. The pretrained ResNet-18 model has a total of 512 input features and 1000 output features, however in this particular case only 3 classes

are outputted i.e., Covid-19, Viral Pneumonia and Normal. Hence, the FC layer is modified to have 3 output features.

The hyper parameters used for training the model are as mentioned in Table 2.

Table 2. Hyper-parameters used for the CNN training

Hyper-parameters	Values
Learning Rate	0.01
Loss Function	Cross Entropy Loss
Batch Size (training)	3917
Batch Size (testing)	1000
Optimizer	Adam
Number of Epochs	1
Steps per Epoch	340

The hyperparameters were decided upon by continuous experimentation and tuning. Cross entropy was decided as the loss function since the model tries to solve a multi classification problem. Adam is an efficient optimizer that proves to be useful in handling sparse gradients, and was chosen over other gradient based optimizers such as SGD and adagrad since it produced greater training accuracy on the ResNet-18 model. The learning rate of adam is 0.01, and this default rate is employed in training the model.

The ResNet18 model is initially set to training phase and gradients are first set to zero before starting to perform backpropragation. Loss is generated from the network parameters and the gradients are computed. After this, loss is optimized by updating the parameters based on the current gradient values. This process is iterated over all images in the train set. At every $20n^{th}$ iteration, the model is set to evaluation phase where the validation loss and accuracy are computed. The model is trained until the validation loss stops dropping for several consecutive iterations i.e., after one epoch of 340 steps which witnessed a minimum validation loss of 0.12.

3.5 Data Poisoning

An adversarial machine learning approach of poisoning the dataset was applied in misguiding the model by providing malicious input. In this paper, the data poisoning approach of data manipulation was employed to manipulate the train data and cause the model to make false predictions. Out of the 7940 covid examples in the train data, 3400 of the images were overlaid with normal chest X-ray images which resulted in manipulated images as shown in Fig. 5. The overlaying of images was performed using Python Imaging Library (PIL) modules in a Jupyter notebook. The corrupted dataset was then used to train the

ResNet-18 model, keeping the hyperparameters the same as before. Training was performed for a single epoch of 340 steps after which the model was tested using the images in the test dataset. The testing accuracy underwent a significant drop of 9% from 0.9648 to 0.8711 indicating the magnitude of the threat that data poisoning imposes on machine learning models.

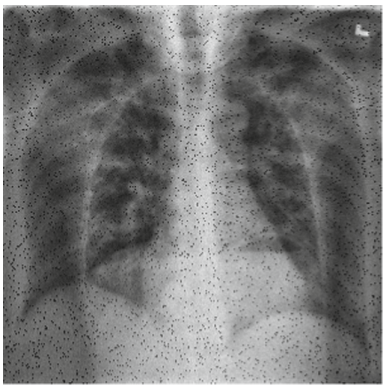


Fig. 5. Manipulated image after overlaying a normal x-ray over a covid-19 x-ray

4 Results

Table 3 shows a comparative analysis of the performance of three different pre-trained networks i.e., ResNet-18, AlexNet and VGG16 after training and testing.

Table 3. Comparative analysis of the performance of 3 models

Value	ResNet-18	AlexNet	VGG16
Validation Accuracy	0.9672	0.9040	0.8748
Validation Loss	0.1236	0.2940	0.3642
Test Accuracy	0.9648	0.9037	0.8552

The classification reports of the three models are as represented in Fig. 6. It can be understood that the ResNet-18 model performs better than the other two pre-trained networks since it delivers optimum precision, recall and f1-score values. The accuracy obtained after testing is 0.9648 which is higher than most state-of-the-art models.

	precision	recall	f1-score	support
0	0.93	0.98	0.96	497
1	0.99	0.97	0.98	474
2	0.97	0.94	0.95	535
accuracy			0.96	1506
macro avg	0.96	0.96	0.96	1506
weighted avg	0.96	0.96	0.96	1506

(a) ResNet-18

	precision	recall	f1-score	support
0	0.84	0.96	0.89	525
1	0.96	0.93	0.94	486
2	0.94	0.83	0.88	495
accuracy			0.90	1506
macro avg	0.91	0.90	0.90	1506
weighted avg	0.91	0.90	0.90	1506

(b) AlexNet

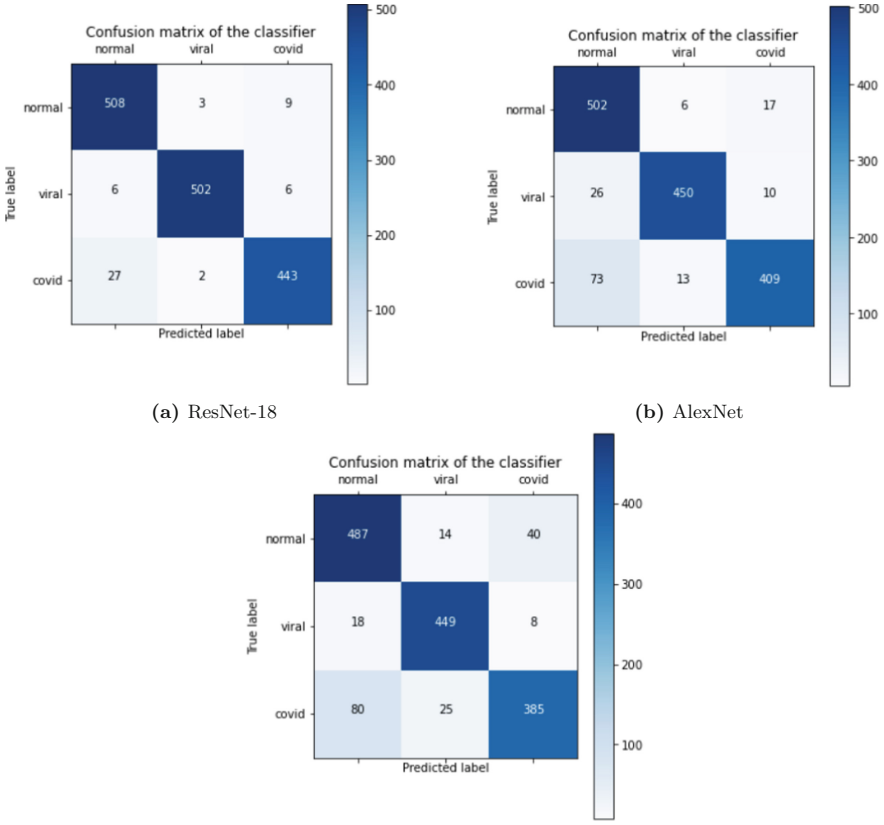
	precision	recall	f1-score	support
0	0.78	0.88	0.82	487
1	0.91	0.95	0.93	491
2	0.89	0.75	0.81	528
accuracy			0.86	1506
macro avg	0.86	0.86	0.86	1506
weighted avg	0.86	0.86	0.85	1506

(c) VGG16

Fig. 6. Classification report

The confusion matrices are plotted as shown in Fig. 7. Testing was performed by selecting a pool of 1500 images (approximately 500 from each of the classes) arbitrarily from the test dataset comprising 6000 images (2000 from each class). The colour scale in the figures suggest that the darker regions depict 400+ images, the lighter regions less than 100 images, while the mid-range colours on the scale denote 100–400 images.

Hence, it can be inferred from the confusion matrix of the ResNet-18 model, that over 450 images from each of the classes were correctly classified while the misclassifications were extremely minimal (less than 30 images per class). Furthermore, it is understood that the AlexNet model does reasonably well in correctly classifying images i.e., above 400 images per class, followed by VGG16 which correctly classifies above 350 images per class. The misclassification rates are considerable low in both the models. i.e., less than 100 images per class.

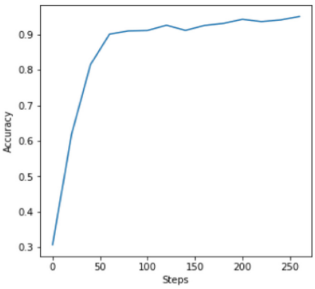


(a) ResNet-18 **(b)** AlexNet

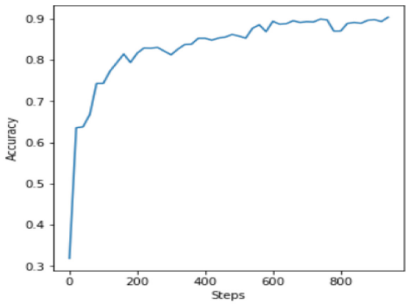
(c) VGG16

Fig. 7. Confusion matrix

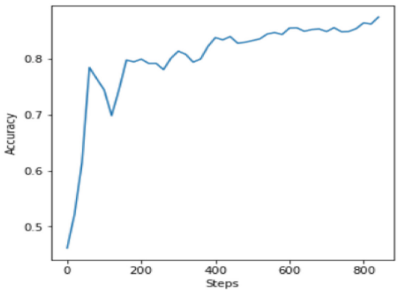
The model accuracy curves are as plotted in Fig. 8. As seen, the curves increase exponentially over time which is desirable during training. The accuracy curves plotted here are on the validation data i.e. steps per epoch vs validation accuracy. It can be inferred that the ResNet-18 curve reaches a maximum validation accuracy of 0.9672, followed by 0.9040 for AlexNet and 0.8748 for VGG16. It can also be noted that the increase in accuracy is seemingly smooth for the ResNet-18 curve, while the other two curves are considerably erratic in their progression.



(a) ResNet-18



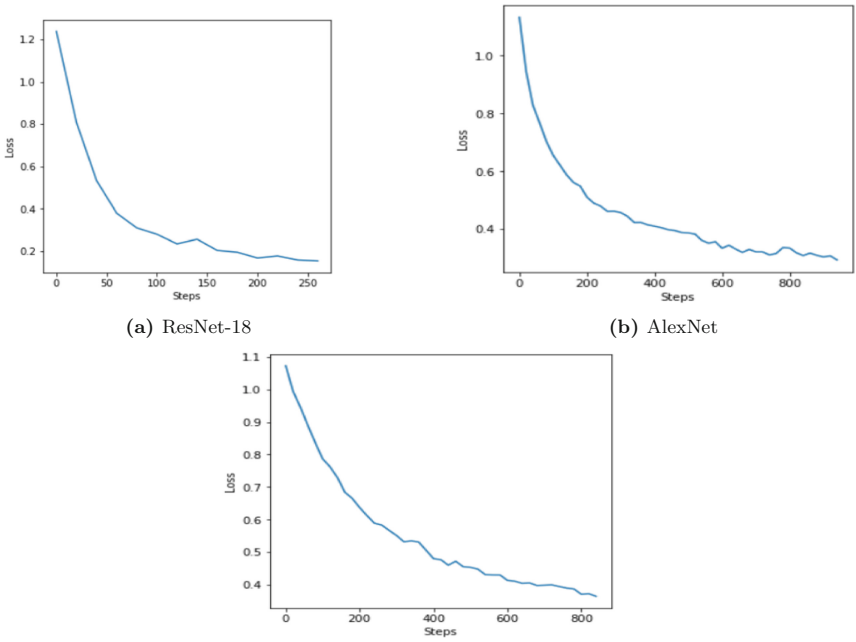
(b) AlexNet



(c) VGG16

Fig. 8. Accuracy curves

The model loss curves are as plotted in Fig. 9. Here, the curves decrease exponentially over time which is ideal while training the model. The loss curves plotted here are on the validation data i.e. steps per epoch vs validation loss. It can inferred that the ResNet-18 curve touches a minimum validation loss of 0.1236, followed by 0.2940 for AlexNet and 0.3642 for VGG16. The curves show a reasonably smooth decline for all three models.



(a) ResNet-18 **(b)** AlexNet

(c) VGG16

Fig. 9. Loss curves

Figure 10 demonstrates the classification report of the ResNet-18 model after data poisoning. As seen there is a significant drop in the accuracy rate from 0.9648 (before poisoning) to 0.8711.

	precision	recall	f1-score	support
0	0.96	0.94	0.95	494
1	0.78	0.97	0.86	507
2	0.92	0.70	0.79	505
accuracy			0.87	1506
macro avg	0.88	0.87	0.87	1506
weighted avg	0.88	0.87	0.87	1506

Fig. 10. Classification report of ResNet-18 after data poisoning

Table 4 summarizes a comparative analysis of the performance of the ResNet-18 model before and after data poisoning.

Table 4. Comparative analysis of the ResNet-18 model before and after data poisoning.

Value	Before Data Poisoning	After Data Poisoning
Train data	8192 normal examples, 7366 viral examples, 7940 covid examples	8192 normal examples, 7366 viral examples, 7933 covid examples 3400 covid images were overlaid with normal images as part of a data manipulation poisoning attack
Test data	2000 normal examples, 2000 viral examples, 2000 covid examples.	2000 normal examples, 2000 viral examples, 2000 covid examples
Validation Accuracy	0.9672	0.8818
Validation Loss	0.1236	0.2973
Test Accuracy	0.9648	0.8711

5 Conclusion and Future Work

Detection of COVID-19 using chest x-rays can help in advancing its diagnosis. This paper details the approach that has been employed in achieving the classification task, and also rationalizes the different implementation choices made such as selection of hyper parameters and model architecture. The ResNet-18 model detailed here shows an optimum accuracy result of 0.9648 which is higher than most state-of-the-art models. A comparative analysis of the model's performance with other pre-trained networks proved that the ResNet-18 model delivered better classification results. The results obtained from data poisoning (9% drop in accuracy from 0.9648 to 0.8711) indicate that it poses a realistic threat to machine learning models, thereby making it extremely easy for an adversary to insert benign looking files and corrupt machine learning models. By expanding the dataset and fine tuning the parameters used, the classification accuracy can be increased even further, which is a scope for future study. The implementation of a Covid-19 CT-severity score in the model is another area of future analysis.

References

1. Phankokkruad, M.: COVID-19 pneumonia detection in chest x-ray images using transfer learning of convolutional neural networks. In: MDSIT 2020: Proceedings of the 3rd International Conference on Data Science and Information Technology, pp. 147–152 (July 2020)
2. Yee, S.L.K., Raymond, W.J.K.: Pneumonia diagnosis using chest x-ray images and machine learning. In: ICBET 2020: Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology, pp. 101–105 (September 2020)

3. Siddiqi, R.: Automated pneumonia diagnosis using a customized sequential convolutional neural network. In: ICDLT 2019: Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, pp. 64–70 (July 2019)
4. Yu, H., Xu, X., Zhao, Z., Li, D.: YU-net lung segment image preprocess methods used for common chest diseases prediction. In: ICMLT 2020: Proceedings of the 2020 5th International Conference on Machine Learning Technologies, pp. 68–71 (June 2020)
5. Makris, A., Kontopoulos, I., Tserpes, K.: COVID-19 detection from chest x-ray images using deep learning and convolutional neural networks. In: SETN 2020: 11th Hellenic Conference on Artificial Intelligence, pp. 60–66 (September 2020)
6. Lin, T.C., Lee, H.C.: COVID-19 chest radiography images analysis based on integration of image preprocess, guided grad-cam, machine learning and risk management. In: ICMHI 2020: Proceedings of the 4th International Conference on Medical and Health Informatics, pp. 281–288 (August 2020)
7. Ke, Q., et al.: A neuro-heuristic approach for recognition of lung diseases from x-ray images. Expert Syst. Appl. **126**, 218–232 (2019)
8. Woźniak, M., Połap, D.: Bio-inspired methods modeled for respiratory disease detection from medical images. Expert Syst. Appl. **41**, 69–96 (2018)
9. Capizzi, G., Sciuto, G.L., Napoli, C., Połap, D., Woźniak, M.: Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bioinspired reinforcement learning. IEEE Trans. Fuzzy Syst. **28**(6), 1178–1189 (2020). https://doi.org/10.1109/TFUZZ.2019.2952831
10. Kaggle's COVID-19 Radiography Database. https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
11. IEEE8023's Covid ChestXray Dataset. https://github.com/ieee8023/covid-chestxray-dataset
12. Kaggle's COVID-19 X rays. https://www.kaggle.com/andrewmvd/convid19-X-rays