ISSN 2976-1190 (Online)
Journal of Engineering and Sciences
Vol. 1 Issue 1

COVID-19 Detection from Chest X-RAY Images Using Ensemble of Deep Learning Algorithms

Ghanashyam Subedi[1], Bidur Devkota[2]
[1]*Department of Software Engineering,Gandaki College of Engineering and Science, Pokhara University, Nepal*
[2] *Department of Software Engineering,Gandaki College of Engineering and Science, Pokhara University, Nepal*

Abstract

Coronavirus disease (COVID-19) is an illness caused by a newly discovered severe acute respiratory syndrome coronavirus 2(SARS-CoV-2). Medical reports reveal that a COVID-19 contaminated individual may experience respiratory symptoms such as coughing, sore throat, shortness of breath and in severe cases affects both lungs. Therefore, the lung can be an important interior organ to analyze the seriousness of COVID-19 infection using X-Ray images of the chest. As there are an enormous number of patients in medical clinics, it would be tedious and hard to analyze stacks of X-ray images, so it tends to be extremely helpful to build up an Artificial Intelligence network that takes care of this work automatically. In this research, Bagging ensemble with majority voting of DenseNet121, ResNet50 and EfficientNetB0 is used. In terms of dataset, freely published X-ray images (10192 Normal,3616 COVID-19,1345 Viral pneumonia and 6012 Lung Opacity) is used. Models trained from scratch performed better than models trained on ImageNet dataset because ImageNet data do not contain X-rays and the features learned from these data were not performing well in X-rays classifications. After fine-tuning DenseNet121 and ResNet50 on models trained from scratch, we received the accuracy of 78 %

Keywords: Ensemble; Convolutional Neural networks; COVID-19; Deep learning

1. Introduction

In December 2019, the novel COVID-19 had begun to spread in China, then in other numerous nations around the globe [10]. The World Health Organization has declared the outbreak a pandemic [8]. From the beginning up until now, the virus has infected at least 40 million people and killed around 1.2 million patients [5]. It is caused by a virus known as SARS-CoV-2 [14]. The virus spreads essentially through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. The vast majority of people infected with the COVID-19 infection will encounter gentle to moderate respiratory sickness and recuperate without requiring exceptional treatment [4].

*Corresponding author. Tel.: +977- 986-3490-195
E-mail address: subedi.ghanshyam@gmail.com

Respiratory issues are one of the main indications of COVID-19, which can be recognized by the X-ray imaging of the chest [9]. X-ray of the chest can also show an ailment that has mild symptoms, so analyzing these images can well recognize the presence of the ailment in suspicious people and even without the symptoms at first [15]. Utilization of these images of the chest have a moderately decent capacity to identify the illness, in the absence of common symptoms.

2. Materials and Method

Ensembles are predictive models that aggregate predictions from two or more different models to create a best predictive model. By combining various models, ensemble learning improves results. Many researchers have shown the superior performance of ensemble learning in classification tasks in their research [17]. The idea of the ensemble learning approach is that each learning model compensates for the deficiencies of other learning models, resulting in superior results [2].Therefore, ensemble learning models are used to detect COVID-19 from X-rays.

2.1 Dataset

COVID-19 Radiography Database [3],[12].
This data set for normal,COVID-19 and other lung infections was made available in phases. The scientists published 219 chest X-rays labeled" COVID-19",1341 "normal" and 1345 "viral pneumonia" in the first release. 1200 CXR images of COVID-19 class were included in the initial upgrade. The database was expanded in the second update to include 1345 cases of viral pneumonia, 3616 COVID19 positive cases,10192 normal and 6012 lung opacity images.

Table 1. Summary of data

X-ray Type	Number of Images
COVID-19	3616
Normal	10192
Lung Opacity	6012
Viral Pneumonia	1345

2.2 Methodology

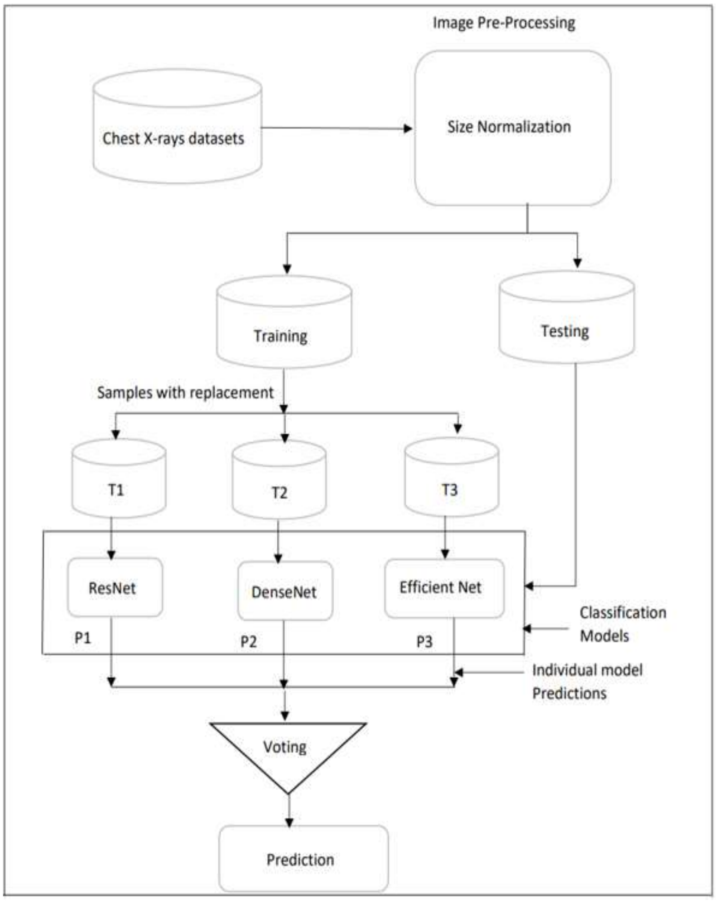

Figure 1: Research Methodology

In order to use an ensemble of learners, be able to combine the learners and get better accuracy, a set of learners which make independent errors are required. In order to force the learners to make independent errors, different data samples can be fed into it and output can be combined by voting method. Bagging Ensemble is one of the most popular ensemble techniques [11]. Figure 1 , represents the methodology used in this research.

2.2.1 Size Normalization

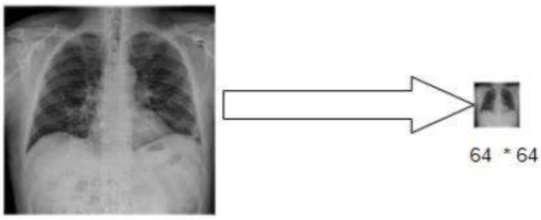

Figure 2: Size Normalization

As shown in Figure 2, Image of different sizes are converted to uniform size of 64 * 64 because deep learning models train faster on small images [13].

2.2.1 Bootstrapping

Bootstrapping is a statistical procedure that generates samples of size K from a dataset of size N by randomly drawing with replacement K observations. These bootstrapped samples are used to train the models.

From the original dataset, many subsets are produced by selecting samples with replacement. These subsets of data are fed into the base models and the final predictions are performed by combining the base model's predictions.

2.2.2 Classification Models

CNNS have gained popularity because of their enhanced picture categorization capabilities. Pretrained models such as ResNet50, DenseNet121 and EfficientNetB0 are used. These models were trained using one million photos from the various class of ImageNet dataset. Because of this, features extracting capabilities of these models can be used on a smaller dataset such as in this study.

ResNet was proposed by researchers from Microsoft Research in 2015 who was the winner of the ImageNet competition winner of 2015. This solved the issue of vanishing gradient by implementing a technique known as skip connections by skipping one or more layers [6]. By skipping some levels in between, the skip connection links activations of one layer to those of other layers. This creates a block of residue. These leftover building blocks are stacked to form ResNet. This obtained accuracy of 77.15% in 2015 [1].

In a conventional feed-forward convolutional neural network (CNN), the first convolutional layer, which receives the input, is the only one that receives the output of the convolutional layer before it. This convolutional layer then creates an output feature map, which is then passed on to the subsequent convolutional layer. However, when the CNN's layer count climbs, or as the layers become deeper, the vanishing gradient problem emerges. DenseNet alleviate this issue by altering the typical CNN architecture and streamlining the connectivity between layers. Each layer in a DenseNet architecture is directly connected to every other layer. It requires few parameters to traditional neural networks since it ignores redundant feature maps from learning [7].

Convolutional Neural Networks are often constructed at a set resource budget, then scaled up for higher accuracy if more resources are available. EfficientNet [16] investigated model scaling and discovered that cautiously balancing the network's depth, width, and resolution can improve performance

3. Results and Discussion

We performed 4 experiments to find out how our ensemble of the CNN models perform. Datasets were divided as 70 % for training, 10% for validation and 20% for testing. All of the experiments were performed by setting following parameters value as below:

Table 2. Parameter values

Name	Value
Epoch	50
Learning rate	0.001
Batch size	32
Optimizer	Adam
Loss function	Cross Entropy

Optimizers alter the neural network's weights and learning rate with the aim to minimize losses. Found Adam to be robust and suitable for various range of non-convex optimization issues in machine learning [18]. A loss function compares the predicted and target output values. Cross Entropy outperforms other loss functions by using entropy to quantify the differences between the prediction distribution and the real distribution [19]. When the model's parameters are changed, the learning rate, a hyperparameter, indicates how much the model should alter in relation to the mistake. Learning rate of 0.001 was set as it performs better in [21]. The batch size, the quantity of samples a model considers before deciding which way to send each weight. A batch size of 32 is an acceptable default value [20] and we tried experimenting with lower and higher batch size of 13 and 64, we received better result when batch size was set as 32. Epoch is one full iteration of the algorithm using the training dataset. For consistency of the results number of epochs was set to 50.

Table3 shows the results of an ensemble of models that were trained on ImageNet data.

Table 3. Models trained on ImageNet data

Models	Accuracy	F1 score	Precision	Recall
DenseNet121	0.46	0.34	0.82	0.21
ResNet50	0.22	0.28	0.89	0.16
EfficientNetB0	0.42	0.30	0.74	0.19
Bagging	0.32	0.31	0.91	0.19

Clearly, results were not as expected, the reason being because models were trained on ImageNet data that do not include X-rays. Therefore, models were trained on X-rays used in this research from scratch. Table 4 shows the results obtained from it.

Table 4. Models build from scratch and trained on X-rays data

Models	Accuracy	F1 score	Precision	Recall
DenseNet121	0.77	0.59	0.98	0.42
ResNet50	0.754	0.57	0.95	0.40
EfficientNetB0	0.713	0.538	0.97	0.37
Bagging	0.76	0.59	0.98	0.42

Clearly, when models are built from scratch and trained on X-rays data, Accuracy increases.

Table 5 and Table 6 shows layers added on top DenseNet121 and ResNet50 respectively to fine-tune them. Table 7 indicates the result we received after DenseNet121 and ResNet50 were fine-tuned on models that were built from scratch and trained on X-rays data.

Table 5. Layers added on top of DenseNet121 to fine-tune

Layer of 1024 inputs and 512 outputs
ReLU
Dropout of 0.25
Layer of 512 inputs and 256 outputs
ReLU
Dropout 0.25
Layer of 256 inputs and 2 outputs
Sigmoid activation

Table 6. Layers added on top of ResNet50 to fine-tune

Layer of 2048 inputs and 512 outputs
ReLU
Dropout of 0.25
Layer of 512 inputs and 2 outputs

Table 7. Fine-tuned Models build from scratch and trained on X-rays data

Models	Accuracy	F1 score	Precision	Recall
DenseNet121	0.79	0.62	0.96	0.45
ResNet50	0.80	0.62	0.93	0.46
EfficientNetB0	0.713	0.538	0.97	0.37
Bagging	0.78	0.61	0.98	0.44

With fine tune models, accuracy of the individual models further increased and thus improved the accuracy of the Bagging ensemble.

We also compare the result of Bagging ensemble with Stacking ensemble and below is the result.

Table 6. Stacking ensemble on fine-tuned models build from scratch and trained on X-rays data

Models	Accuracy	F1 score	Precision	Recall
Stacking Ensemble	0.634	0.440	0.842	0.298

For stacking, SVM was used as a Meta classifier.

Transfer learning allows you to use feature representations from a model that has already been trained rather than having to create a new model from start. Weights derived from the models can be applied to other computer vision tasks. However, in this case since the models were trained on ImageNet data that do not contains X-rays, weights derived from this were not useful in classifying X-rays.

Therefore, models trained from scratch performed well and thus their ensemble. In addition, since models were created to classify 1000 class categories of ImageNet dataset, extra layers were added on top of the model to fine-tune the models. Fine-tuned DenseNet121 and ResNet50 improved the performance of individual models and thus improved Bagging ensemble performance. Bagging ensemble has a precision of 0.98, which means that 98% of the time it correctly identifies an X-ray as COVID-19. Recall of 0.44 means that it accurately predicts 44% of all COVID-19 X-rays.

4. Conclusions

In this paper, we created an ensemble model of DenseNet121, ResNet50 and EfficientNetB0 and used majority voting to predict the final output. Accuracy of the ensemble models increased drastically when trained from scratch with X-ray data as compared to performing transfer learning on ImageNet data. After, we performed tuning on DenseNet121 and ResNet50 that were created from scratch, accuracy of the ensemble model increased by 2% . Comparing Bagging ensemble with Stacking ensemble on fine tune models, accuracy of Bagging was higher than Stacking. Bagging is effective on unobserved data because it reduces variance of base models. Performance indicators such as accuracy, precision, recall and F1-Score are used to analyze the experimental efforts. A better statistic would be the F1 score for this experiment where data are unbalanced. It represents the harmonic mean of recall and precision.

4.1 Limitations

COVID-19 X-rays data are less compared to Non-COVID-19 X-rays.

4.2 Future work

As a next step, we are going to increase the COVID-19 X-rays data set by collecting from freely available sources and fine tune the EfficientNetB0.

In addition, dataset will be balanced by applying augmentation technique in COVID-19 X-rays images.

References

[1] Nikolas Adaloglou. Best deep CNN architectures and their principles: from AlexNet to EfficientNet. Jan. 2021. URL: https : / / theaisummer . com / cnn -architectures/.

[2] Wafa Boukellouz and Abdelouahab Moussaoui. "Magnetic resonance-driven pseudo CT image using patch-based multi-modal feature extraction and ensemble learning with stacked generalisation". In: Journal of King Saud University-Computer and Information Sciences (2019).

[3] Muhammad EH Chowdhury et al. "Can AI help in screening viral and COVID-19 pneumonia?" In: arXiv preprint arXiv:2003.13145 (2020).

[4] Coronavirus. URL: https://www.who.int/health-topics/coronavirus#tab=tab_1 .

[5] Coronavirus Cases: Oct. 2020. URL: https : / / www . worldometers .info/coronavirus/.

[6] Kaiming He et al. "Deep residual learning for image recognition". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016,pp. 70–778.

[7] Gao Huang et al. "Densely connected convolutional networks". In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017,pp. 4700–4708.

[8] HM Iqbal et al. "The emergence of novel-coronavirus and its replication cycle:an overview". In: J Pure Appl Microbiol 14.1 (2020), pp. 13–16.

[9] Fang Jiang et al. "Review of the clinical characteristics of coronavirus disease 2019 (COVID-19)". In: Journal of general internal medicine (2020), pp. 1–5.

[10]Shervin Minaee et al. "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning". In: arXiv preprint arXiv:2004.09363 (2020).

[11]Li Quo and Samia Boukir. "Building an ensemble classifier using ensemble margin. Application to image classification". In: 2017 IEEE International Conference on Image Processing (ICIP). IEEE. 2017, pp. 4492–4496.

[12]Tawsifur Rahman et al. "Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection using Chest X-rays Images". In: Computers in Biology and Medicine (2021), p. 104319.

[13]Sergio Saponara and Abdussalam Elhanashi. "Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance". In: International Conference on Applications in Electronics Pervading Industry, Environment and Society. Springer. 2022, pp. 10–17.

[14]Sibylle Bernard Stoecklin et al. "First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020".In: Eurosurveillance 25.6 (2020), p. 2000094.

[15]Dan Sun et al. "Clinical features of severe pediatric patients with coronavirus disease 2019 inWuhan: a single center's observational study". In: World Journal of Pediatrics (2020), pp. 1–9.

[16]Mingxing Tan and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: International Conference on Machine Learning.PMLR. 2019, pp. 6105–6114.

[17]Yun Yang. Temporal Data Mining via Unsupervised Ensemble Learning. Elsevier,2016.

[18]D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 2014, 2014.

[19]Y. Zhou, X. Wang, M. Zhang, J. Zhu, R. Zheng, and Q. Wu,

Mpce: A maximum probability based cross entropy loss function for neural network classification, IEEE Access, vol. 7 2019, pp. 146 331-146 341, 2019.
[20]Y. Bengio, Practical recommendations for gradient-based training of deep architectures, in Neural networks: Tricks of the trade, Springer, 2012, pp. 437–478.
[21]M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, Efficient pneumonia detection in chest xray images using deep transfer learning, Diagnostics, vol. 10, no. 6 2020,p. 417, 2020.