Deep Learning

880008-M-6

Assignment

Using Deep Learning to Perform Multi-Class Classification on the

Covid19 Chest X-ray Dataset

Report by:

Andras Tuu (2107756)

Group Number:

4

Group Members:

Isabel van Dam

Andras Tuu  
Hai Nam Nguyen

March 2023

1. Problem Definition

Computer vision is important for medical data because it can provide a more accurate and efficient way to make diagnoses, detect abnormalities, and guide treatment decisions. It can also be used to detect patterns in large datasets, helping healthcare providers identify trends and changes over time. Additionally, computer vision can help to automate the analysis of medical data, reducing the time and cost of manual review processes. It can also help to diagnose patients and inform treatment decisions quickly and accurately. It can also help to identify areas of high risk that can be targeted for additional testing and preventive measures. Additionally, it can reduce the burden on healthcare providers by eliminating the need to manually review X-ray images and manually diagnose patients, freeing up valuable resources.

There have been very interesting trials for example MediPredict, which is a Deep Learning based method for predicting based on past and recent medical data the risk of having or getting specific diseases, and how to prevent them. (https://medipredict.com)

1. Dataset Preprocessing

The following part is about preprocessing of the dataset and the Exploratory Data Analysis. The dataset consists of 6392 images, with dimensions 156 x 156. Out of these 6392 cases 2816 have bacterial pneumonia, 127 have COVID-19, 1606 are healthy and 1843 have Viral Pneumonia.

The first preprocessing step is one-hot encoding, we chose this method because it is a simple and efficient way to represent categorical data., and it also helps with the interpretation and analysis. Then we split the images into train, test and validation sets using stratified splits with ratios of 0.2 and random state of 42. In the last part we normalized the data. The training set consists of 4090 images, the validation set consists of 1023 while the test set has 1279 pictures. Lastly, we normalize the data to reduce noise and improve the accuracy of the model. Also, this way the model can predict better.

1. Baseline Model

The model consists of a series of convolutional neural network layers, a max pooling layer, and two dense layers. The first layer of the model is a convolutional neural network with 64 filters and a window size of 3x3. This is followed by another convolutional neural network layer with 32 filters and a window size of 3x3. Both layers use the same input shape as the training data and use the ReLU activation function.

Next, the model uses a max pooling layer with a window size of (2,2). This layer operates on the output of the convolutional layers and reduces the dimensionality of the output. Following this, the model uses two additional convolutional layers with the same parameters as the first two layers.

After the convolutional layers, the model flattens the output and passes it through two dense layers with 32 neurons and a ReLU activation function. Finally, the model uses a dense layer with 4 neurons and a softmax activation function to output the predicted class.

The model is trained using the Adam optimizer, categorical cross entropy loss and accuracy metric. A model checkpoint is used to save the weights of the epoch with the highest validation accuracy.

Text

Description automatically generatedText

Description automatically generated

(a) Validation accuracy, F1, Sensitivity, Specificity (b) Test accuracy, F1, Sensitivity, Specificity

Figure 1: Confusion Matrixes

The results showed that the model had a validation accuracy of 0.781, which is a good result for a baseline model. The validation and test sets both had an accuracy of 0.775 and 0.764. The F1-scores for each class were comparable across both sets, with class 0 having 0.804 and 0.802, class 1 having 0.077 and 0.061, class 2 having 0.892 and 0.877, and class 3 having 0.657 and 0.632.

Overall, the baseline deep learning model was effective in predicting based on X-ray of lungs, with a good accuracy and F1-score. However, the sensitivity and specificity of the model needs improvement, the low specificity score in the case of COVID-19 and No Pneumonia means that the it has a low false positive rate, but a high false negative rate.

Increasing cross entropy loss on the validation set and decreasing on the training set means that the model is overfitting on the training set, as it is not generalizing well to the unseen data from the validation set. The classification accuracy increasing more on the training dataset also point to the model overfitting.

Chart, line chart

Description automatically generated

Figure 2: Cross Entropy Loss and Classification Accuracy

The confusion matrixes show that 77 cases on the validation dataset were assigned as healthy while they should have been classified as having Viral Pneumonia and 82 cases were assigned as healthy while they had Viral Pneumonia. Also, most cases for COVID-19 had been assigned to healthy. The test confusion dataset follows a similar pattern. The classification of COVID-19, and false positives and negatives in the Viral Pneumonia and healthy category are the main problems, with the base model.

A picture containing chart

Description automatically generatedA picture containing chart

Description automatically generated

(a) Confusion Matrix validation (b) Confusion Matrix test

Figure 3: Examples of confusion matrixes

As it is visible on the ROC curves, the model is performing the worst on classifying the healthy class, as I have stated it before. Furthermore, it is visible, that the COVID-19 do not have enough data as the curve is too angular.

Chart, diagram

Description automatically generatedDiagram

Description automatically generated

(a) ROC curve validation (b) ROC curve test

Figure 4: Examples of ROC curves

1. Improved (Fine-tuned) Model

We tried different augmentation methods, rotating, shifting the height and width, flipping vertically, horizontally and shearing the optimal solution was to use rotating zooming and vertically flipping the images. With values 0.3, 0.1 for rotating and zooming.

In the fine-tuned model we have increased the number of convolutional layers to 6, we have also introduced 2 dropout layers, 4 pooling layers and 3 dense layers. We have changed up the filters as well, starting with 32, then increasing to 64,128,256, with a pooling layer in-between. Before going to the dense layers, we have a flattening layer. Then 3 dense layers, the first 2 with size of 64 while the last (output) layer has a size of 4, because of the 4 classes. We have tried different parameters on the dropout layers, L2 regularization, and the Adam optimizer, while also added early stopping, with the metric validation accuracy. In the end the best parameters were 0.2 for dropout layers, 0.0001 for L2 regularization, the epochs we increased to 100 and changed the batch size to 32. Most of these methods and tuning was based on the research of (Meedeniya et al.,2022),(Ahmed et al.,2021)

The most important paper we have found on loss functions, VGG model, optimizers and how to counter overfitting is (Meedeniya et al.,2022). We tried regularization, early stopping and feature removing. We kept regularization and early stopping, because it helped our model achieve better validation and test accuracy, while also reducing overfitting

Ahmed et al. have presented a CNN model with five convolutional layers, each followed by batch normalization and max-pooling layers, and a dropout where the final layer is fully connected. We tried dropout layers and max-pooling layers based on this article, and kept them for our final model, these methods also helped with overfitting and getting better accuracy on validation and test set.

Table

Description automatically generatedTable

Description automatically generated

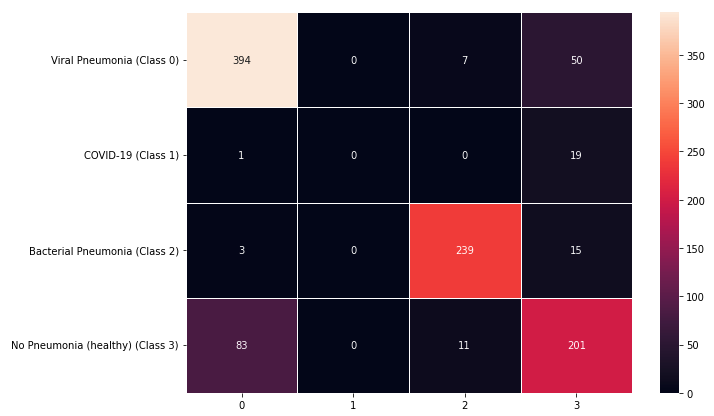
(a) Validation accuracy, F1, Sensitivity, Specificity (b) Test accuracy, F1, Sensitivity, Specificity

Figure 5: Confusion Matrixes

Chart, line chart

Description automatically generated

Figure 6: Cross Entropy Loss and Classification Accuracy

A picture containing chart

Description automatically generated

(a) Confusion Matrix validation (b) Confusion Matrix test

Figure 7: Examples of confusion matrixes

Diagram

Description automatically generated with medium confidenceDiagram

Description automatically generated

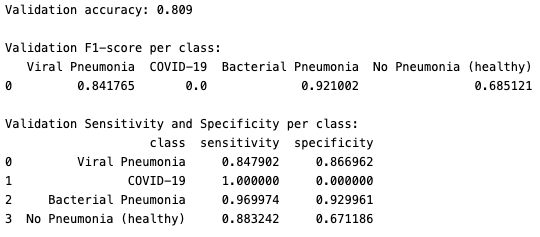
(a) ROC curve validation (b) ROC curve test

Figure 8: Examples of ROC curves

With the tuned model and augmentation, we managed the problems that were overfitting and the cross-entropy loss and classification accuracy improved, however the model did not solve the problems with the classification of COVID-19, and false positives and negatives in the Viral Pneumonia and healthy category are the still the main problems with the tuned model, however the number of false negatives have decreased. The ROC curves show us that the model generally got better.

1. Transfer Learning Model

As for the transfer learning model, we have tried the Resnet 50 model and the VGG16 model, and the VGG16 model has performed better on both validation and test accuracy. We had the same overfitting problem at first with this model but the dropout layers and the l2 regularization, we have managed to get the model to work. We have used a flattening layer, two dense layers with 64 filters, and 2 dropout layers with values 0.2. We used the Adam regularizer with learning rate 0.01 and categorical cross entropy as our loss function. We have also used early stopping and model checkpoint. We used 50 epochs and batch size of 64, we used the same augmentation process as we have used with the tuned model.

Table

Description automatically generated

(a) Validation accuracy, F1, Sensitivity, Specificity (b) Test accuracy, F1, Sensitivity, Specificity

Figure 9: Confusion Matrixes

Chart, line chart

Description automatically generated

Figure 10: Cross Entropy Loss and Classification Accuracy

A picture containing chart

Description automatically generatedA picture containing chart

Description automatically generated

(a) Confusion Matrix validation (b) Confusion Matrix test

Figure 11: Examples of confusion matrixes

Diagram

Description automatically generated with medium confidenceDiagram

Description automatically generated with low confidence

(a) ROC curve validation (b) ROC curve test

Figure 12: Examples of ROC curves

The model achieved similar accuracy and sensitivity and specificity values as the tuned model, the model as it can be seen on the cross entropy and classification accuracy curves is overfitting a bit on the training data, unfortunately the computing power that we had was not enough to make further changes in this model. We could have tried more different regularization constraints, different dropout values or different l2 coefficients, we could have also used different optimizers. Also, for most of these problems we could have used keras tuner, but unfortunately, we did not have the computational power for that.

1. Discussion

The models generally performed well, the main problem we were facing is overfitting and the classification of the harder to classify images (COVID-19, Healthy, Viral Pneumonia). We managed to handle the problem of overfitting, and we could have gotten better results, if we used keras tuner for the tuning of the parameters, unfortunately we did not have the processing power for that. The transfer learning method in my opinion has the highest potential, with the right tuning. In the articles mentioned below there are more methods and architectures that have the potential to get better results. The biggest limitation for us was computing power, and running out of free GPU usage on google collab. Regardless, I have learned a lot and enjoyed the project thoroughly.

1. References

Ahmed F., Bukhari S.A.C., Keshtkar F. A deep learning approach for COVID-19 8 viral pneumonia screening with X-Ray images. Digit. Gov.: Res. Pract. 2021;2(2) doi: 10.1145/3431804.

Meedeniya D, Kumarasinghe H, Kolonne S, Fernando C, Díez IT, Marques G. Chest X-ray analysis empowered with deep learning: A systematic review. Appl Soft Comput. 2022 Sep;126:109319. doi: 10.1016/j.asoc.2022.109319. Epub 2022 Jul 18. PMID: 36034154; PMCID: PMC9393235.