PNEUMONIA LUNGS DISEASE PREDICTION WITH DEEP LEARNING

SRUJANA KONDA
dept. of Computer Science
700740479

REVANTH BHARADWAJ TARIMELA ROOPAVATARAM
dept. of Computer Science
700743276

SAI PAVAN REDDY POGULA
dept. of Computer Science
700741739

TEJASWI REDDY MARRI
dept. of Computer Science
700742730

Abstract- This study focuses on developing a deep learning algorithm to predict Pneumonia, a respiratory disease that affects the lungs' alveoli. The research utilizes patient records with different attributes related to the disease and passes them into the algorithm for analysis. Pneumonia is typically caused by viral or bacterial infections, and the severity of the condition varies, with symptoms including coughing, chest pain, fever, and difficulty breathing. Risk factors for Pneumonia include conditions such as cystic fibrosis, asthma, and a weak immune system.

The diagnosis of Pneumonia is usually based on symptoms and physical examination, with additional tests such as chest X-rays, blood tests, and sputum cultures used to confirm the diagnosis. The study focuses on developing a deep learning algorithm to predict Pneumonia based on the patient's medical reports. The algorithm classifies and divides the data into segments, each containing information on whether the patient has Pneumonia or not. The algorithm is trained to analyze data related to other diseases related to Pneumonia to predict the output accurately.

The research is crucial for hospitals, as the ability to predict Pneumonia accurately can help in diagnosing and treating patients more effectively. In conclusion, this study emphasizes the importance of developing accurate and reliable algorithms for predicting Pneumonia. The research utilizes deep learning algorithms to analyze patient data and predict the presence or absence of the disease accurately. The results of this study can be instrumental in improving patient care and medical diagnosis in hospitals.

Keywords- Pneumonia, Respiratory disease, Deep Learning Algorithm, Patient records, Viral infections, Bacterial infections, Risk factors, Medical reports, Accurate prediction, Patient care.

I. INTRODUCTION

Pneumonia is a common respiratory disease that affects millions of people around the world each year. It is a serious condition that can cause a range of symptoms, including coughing, chest pain, fever, and difficulty breathing. Pneumonia can be caused by a variety of factors, including viruses, bacteria, and other microorganisms.In recent years, there has been a growing interest in using artificial intelligence (AI) and machine learning (ML) to improve the diagnosis and treatment of pneumonia. Deep learning algorithms, in particular, have shown promise in detecting patterns in medical data that can help identify patients with pneumonia.

https://github.com/pavan-reddy-28/NN-DL_Final_project

The use of deep learning algorithms in pneumonia detection is an important development in the field of healthcare. With the help of these algorithms, doctors and healthcare providers can quickly and accurately diagnose pneumonia, which can lead to better outcomes for patients.One of the key advantages of using deep learning algorithms for pneumonia detection is their ability to process large amounts of data quickly and accurately. This can help doctors and healthcare providers make faster and more informed decisions about patient care.In addition to improving the diagnosis and treatment of pneumonia, deep learning algorithms can also help identify risk factors for the disease. This can be particularly useful in developing preventative strategies to reduce the incidence of pneumonia.

There are many different types of deep learning algorithms that can be used in pneumonia detection, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These algorithms are designed to learn patterns in medical data and make predictions based on that data.

To implement a deep learning algorithm for pneumonia detection, a dataset of medical records and chest X-ray images can be used. The data is pre-processed to remove any noise or null values, and then analyzed and visualized for further processing.

Once the dataset has been prepared, a deep learning algorithm is chosen and trained on the data. The accuracy of the algorithm can then be evaluated using a variety of metrics, including precision, recall, and F1 score.

Google Colab Python Tool is a popular platform for developing and running deep learning algorithms for pneumonia detection. It allows developers to work directly in the cloud, with access to a range of machine learning libraries and tools

Overall, the use of deep learning algorithms for pneumonia detection is an exciting development in the field of healthcare. By leveraging the power of AI and machine learning, we can improve the accuracy and speed of pneumonia diagnosis, leading to better outcomes for patients.

In recent years, the development of advanced technologies such as deep learning algorithms has enabled more accurate and efficient detection of pneumonia. Deep learning algorithms are a type of artificial intelligence that can analyze large amounts of data and identify patterns that can be used to make predictions. In the context of pneumonia detection, deep learning algorithms can analyze patient records and chest X-ray images to predict whether a patient has pneumonia or not.

Overall, the development of an accurate and efficient deep learning algorithm for pneumonia detection has the potential to significantly improve the diagnosis and treatment of this common respiratory disease. By detecting pneumonia early and accurately, patients can receive timely treatment and avoid the severe health consequences associated with the disease. Furthermore, the use of advanced technologies such as deep learning algorithms can help to streamline the diagnosis process and reduce the burden on healthcare providers, ultimately improving the quality of care for patients with pneumonia.

II. Motivation

Pneumonia is a respiratory disease that poses a significant health threat to individuals worldwide. With millions of people affected each year, the importance of early detection and accurate diagnosis cannot be overstated. Timely intervention is crucial for effective treatment and improved patient outcomes. However, the lack of experienced radiologists and the delay in reaching a conclusion often result in treatment deferral, leading to severe consequences.

To reduce the mortality rate associated with lung illnesses, it is imperative to detect pneumonia early and provide prompt treatment. Identifying the condition in its early stages allows healthcare providers to plan appropriate treatment strategies, ultimately improving patient outcomes. Accurate and timely diagnosis plays a vital role in managing pneumonia and preventing its progression.

Pneumonia can affect individuals of all age groups, from young children to older adults. Typical symptoms include a combination of a productive or dry cough, chest pain, fever, and difficulty breathing. The severity of these symptoms varies depending on the underlying cause and the overall health of the individual. Recognizing the symptoms early on and promptly seeking medical intervention can prevent the disease from worsening and reduce the risk of complications.

Efficient diagnosis and treatment of pneumonia are crucial in minimizing the morbidity and mortality rates associated with this respiratory disease. Radiology plays a critical role in the detection and diagnosis of pneumonia, as it enables non-invasive visualization of the lungs. However, the shortage of experienced radiologists presents a significant challenge, hindering the timely and accurate diagnosis of pneumonia.

In conclusion, pneumonia is a severe and potentially life-threatening respiratory disease that demands timely and accurate diagnosis for effective treatment and improved patient outcomes. Early detection and precise diagnosis can significantly reduce the mortality rate linked to lung illnesses. Identifying pneumonia in its early stages allows for predicting the severity of the condition and implementing appropriate interventions. The shortage of experienced radiologists further emphasizes the urgent need for developing efficient diagnostic tools to facilitate timely and accurate diagnosis of pneumonia. By addressing these challenges, we can enhance the management of this critical respiratory disease and improve patient care worldwide.

https://github.com/pavan-reddy-28/NN-DL_Final_project

III. Main Contributions & Objectives

Finding new cases of pneumonia in patient data is the goal of pneumonia detection. The dataset including the chest X-ray pictures from the Pneumonia infection will be used in this investigation. The dataset will undergo pre-processing, and the dataset will be cleaned up of the noisy and null value data. The data will then be examined and displayed in preparation for further processing. To produce the forecast, a machine learning method will be used.

Here, the clinical perception and clinical images are combined in a manner inspired by the discovery method of human professionals. A practical convolutional neural network is created by combining clinical data with a requirement-based computation.

The collection also includes a significant amount of noisy data. However, better outcomes can be achieved using feature engineering. Importing libraries and loading data comes first. After that, a basic knowledge of the data, such as its form, will be taken, and a sample will be taken to see if the dataset contains any NULL values. Understanding the data is a crucial stage in every machine learning project or prediction. That there are no NULL values is a positive thing.

Starting with the main segment, it will examine each subsequent segment to understand how it affects the goal segment. We will also carry out preprocessing at the required phase and incorporate designing tasks. The goal of conducting a thorough exploratory analysis is to prepare and clean information for improved Convolutional neural networks, which may demonstrate elite performance and summarized models. Therefore, it should start with decomposing and preparing the dataset for expectation.

Modules:

1) Dataset gathering

2) Cleaning up data

3) Investigative Image Analysis

4) Modeling with convolutional neural networks

5) Document

1) Dataset collection: For training and testing, several patient kinds provided information on the pneumonia disease together with various forms of chest x-ray pictures.

2) Data cleaning: To create a high-quality dataset for additional pruning, the vast dataset must be pre-processed to remove more erroneous and noisy data. The process of cleaning and processing the dataset begins with deleting the null values.

3) Investigative Image Analysis

Exploratory analysis is a technique to thoroughly investigate and comprehend the link between the data and the data itself in order to make the phases of feature engineering and machine learning modeling easy and efficient for prediction. It supports proving our presumptions correct or incorrect. In other words, performing hypothesis testing is helpful.

4) Neural networks with convolutions Modeling

Convolutional neural network modeling aids in identifying the ideal method and hyperparameters for maximizing accuracy. There are two versions of the dataset.

The neural network algorithm is trained using data from 50% of the picture recordings. The remaining 50% of the photos are used for testing, which aids in the process' prediction.

5. Report:

In order to evaluate and provide precise data to the user for prediction, the data is displayed based on the output of the convolutional neural networks algorithm and is mapped with various types of graphs. In order to map the results according to the needs of the user, Marplot libraries are built.

Implementation of Machine Learning models:

The dataset is divided into training and testing data. The split dataset is passed into the different deep learning algorithm models and the accuracy levels were found.

Normalize the data:

Perform a grayscale normalization to reduce the effect of illumination differences. Moreover, the CNN converges faster on [0..1] data than on [0..255]

Resize data for deep learning:

The train and test dataset is resized with the pre-processing of the images for the evolution of the deep learning model.

Data Augmentation:

To avoid the overfitting problem, we need to artificially expand our dataset. We can make your existing dataset even larger. The idea is to alter the training data with small transformations to reproduce the variations.

Approaches that alter the training data in ways that change the array representation while keeping the label the same are known as data augmentation techniques. Some popular augmentations people use are grayscales, horizontal flips, vertical flips, random crops, color jitters, translations, rotations, and much more.

By applying just a couple of these transformations to our training data, we can easily double or triple the number of training examples and create a very robust model.

For the data augmentation:

1. Randomly rotate some training images by 30 degrees
2. Randomly zoom by 20% some training images
3. Randomly shift images horizontally by 10% of the width
4. Randomly shift images vertically by 10% of the height
5. Randomly flip images horizontally.

https://github.com/pavan-reddy-28/NN-DL_Final_project

SAME Padding: it applies padding to the input image so that the input image gets fully covered by the filter and specified stride. It is called SAME because, for stride 1, the output will be the same as the input.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 150, 150, 32)      320

 batch_normalization (BatchN  (None, 150, 150, 32)     128
 ormalization)

 max_pooling2d (MaxPooling2D  (None, 75, 75, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 75, 75, 64)        18496

 dropout (Dropout)           (None, 75, 75, 64)        0

 batch_normalization_1 (Batc  (None, 75, 75, 64)       256
 hNormalization)

 max_pooling2d_1 (MaxPooling  (None, 38, 38, 64)       0
 2D)
```

The above model trains and tests the x-ray chest images with the help of convolutional neural networks.

```
[11] learning_rate_reduction = ReduceLROnPlateau(monitor='val_accuracy', patience = 2, verbose=1,factor=0.3, min_lr=0.000001)
```

```
[12] history = model.fit(datagen.flow(x_train,y_train, batch_size = 32) ,epochs = 12 , validation_data = datagen.flow(x_val, y_val) ,callbacks = [learning_rate_reduct
```

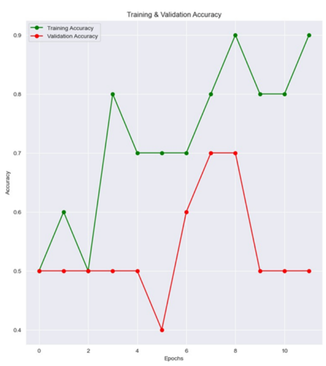
Results:

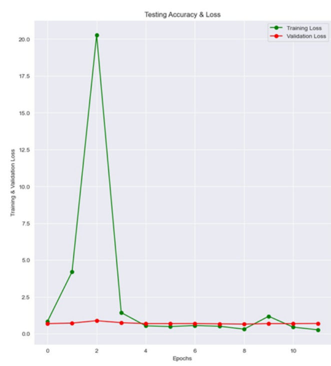
```
Epoch 1/12
1/1 [==============================] - 4s 4s/step - loss: 0.8185 - accuracy: 0.6000 - val_loss: 0.6934 - val_accuracy: 0.5000 - lr: 0.0010
Epoch 2/12
1/1 [==============================] - 1s 1s/step - loss: 12.8000 - accuracy: 0.5000 - val_loss: 0.7177 - val_accuracy: 0.5000 - lr: 0.0010
Epoch 3/12
1/1 [==============================] - ETA: 0s - loss: 14.5727 - accuracy: 0.5000
Epoch 3: ReduceLROnPlateau reducing learning rate to 0.0003000000142492354.
1/1 [==============================] - 1s 1s/step - loss: 14.5727 - accuracy: 0.5000 - val_loss: 0.7555 - val_accuracy: 0.5000 - lr: 0.0010
Epoch 4/12
1/1 [==============================] - 1s 916ms/step - loss: 1.2016 - accuracy: 0.8000 - val_loss: 0.7142 - val_accuracy: 0.5000 - lr: 3.0000e-04
Epoch 5/12
1/1 [==============================] - ETA: 0s - loss: 0.9349 - accuracy: 0.9000
Epoch 5: ReduceLROnPlateau reducing learning rate to 9.000000427477062e-05.
1/1 [==============================] - 1s 992ms/step - loss: 0.9349 - accuracy: 0.9000 - val_loss: 0.6812 - val_accuracy: 0.5000 - lr: 3.0000e-04
Epoch 6/12
1/1 [==============================] - 1s 935ms/step - loss: 0.7690 - accuracy: 0.8000 - val_loss: 0.6850 - val_accuracy: 0.6000 - lr: 9.0000e-05
Epoch 7/12
1/1 [==============================] - 2s 2s/step - loss: 0.0935 - accuracy: 1.0000 - val_loss: 0.6353 - val_accuracy: 0.8000 - lr: 9.0000e-05
Epoch 8/12
1/1 [==============================] - 2s 2s/step - loss: 0.6091 - accuracy: 0.7000 - val_loss: 0.6743 - val_accuracy: 0.5000 - lr: 9.0000e-05
Epoch 9/12
1/1 [==============================] - ETA: 0s - loss: 0.5968 - accuracy: 0.9000
Epoch 9: ReduceLROnPlateau reducing learning rate to 2.700000040931627e-05.
1/1 [==============================] - 1s 843ms/step - loss: 0.5968 - accuracy: 0.9000 - val_loss: 0.7725 - val_accuracy: 0.5000 - lr: 9.0000e-05
Epoch 10/12
1/1 [==============================] - 1s 878ms/step - loss: 0.8542 - accuracy: 0.8000 - val_loss: 0.9045 - val_accuracy: 0.5000 - lr: 2.7000e-05
Epoch 11/12
```

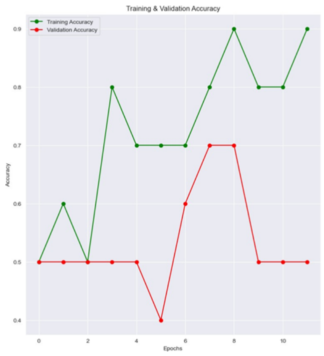
Finding the accuracy of the model:

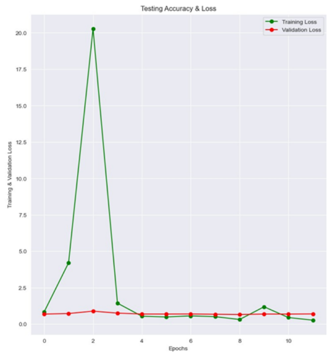
```
[13] print("Loss of the model is - " , model.evaluate(x_test,y_test)[0])
     print("Accuracy of the model is - " , model.evaluate(x_test,y_test)[1]*100 , "%")
```

Results of the accuracy:









The above graph depicts the training accuracy with the validation accuracy of the convolutional neural network.

Predictions

```
predict_x=model.predict(x_test)
predictions=np.argmax(predict_x,axis=1)


predictions = predictions.reshape(1,-1)[0]
predictions[:15]
```

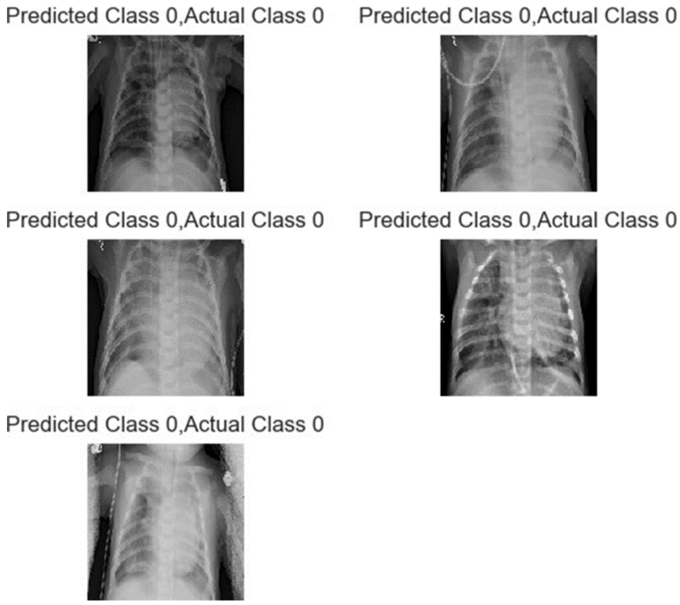
The predictions of the model are evaluated by bypassing the testing and training images of the dataset.

The results of prediction results are displayed to show the precision, recall, f1-score, and support.

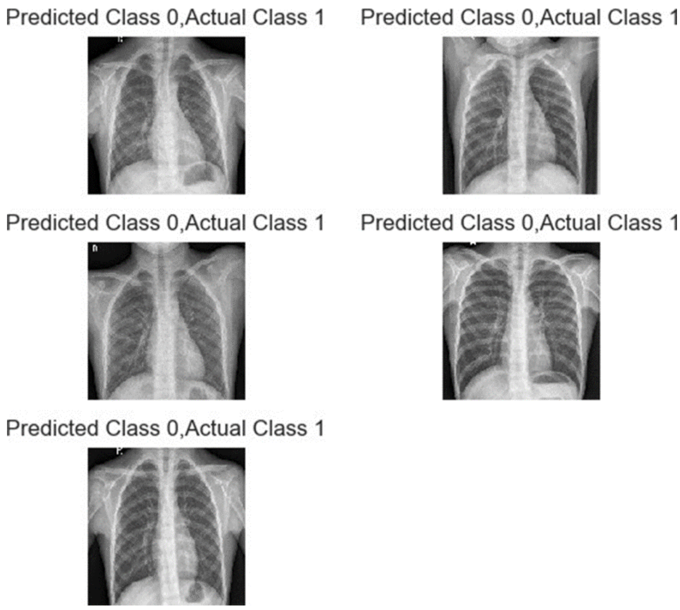
```
[18] print(classification_report(y_test, predictions, target_names = ['Pneumonia (Class 0)','Normal (Class 1)']))

                     precision    recall  f1-score   support

Pneumonia (Class 0)       0.50      1.00      0.67         5
   Normal (Class 1)       0.00      0.00      0.00         5

           accuracy                           0.50        10
          macro avg       0.25      0.50      0.33        10
       weighted avg       0.25      0.50      0.33        10
```

Results of the predicted class:

The predicted class and the actual class x-ray images are displayed. The results also show the incorrectly predicted classes.



The predicted class and the actual class x-ray images are displayed. The results also show the incorrectly predicted classes.



References/Bibliograpy:

[1]. Hao ren2,1∗ aslan b. wong1∗, wanmin lian3∗, weibin cheng2, ying zhang1, jianwei he1, qingfeng liu4, jiasheng yang4, chen zhang5, (member, ieee), kaishun wu1, (member, ieee), haodi zhang1," DOI: 10.1109/ACCESS.2021.3090215, IEEE Access.

[2]. Xuelin Qian, Student Member, IEEE, Huazhu Fu, Senior Member, IEEE, Weiya Shi, Tao Chen, Member, IEEE, Yanwei Fu∗, Member, IEEE, Fei Shan∗, and Xiangyang Xue∗, Member, IEEE "© IEEE 2020.

[3]NCIRD, "https://www.cdc.gov/coronavirus/2019-ncov/lab/guidelinesclinical-specimens.html," 2020.

[4] T. Ai, Z. Yang, H. Hou et al., "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," Radiology, p. 200642, 2020.

[5] M. Chung, A. Bernheim, X. Mei et al., "CT imaging features of 2019 novel coronavirus (2019-nCoV)," Radiology, vol. 295, no. 1, pp. 202– 207, 2020.

[6] Y. Pan and H. Guan, "Imaging changes in patients with 2019-nCov," European Radiology, Feb 2020.

[7] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," Radiology, p. 200432, 2020.

[8] J. P. Kanne, B. P. Little, J. H. Chung, B. M. Elicker, and L. H. Ketai, "Essentials for radiologists on covid-19: an update radiology scientific expert panel," 2020.

[9] M. Liu, W. Zeng, Y. Wen, Y. Zheng, F. Lv, and K. Xiao, "COVID-19 pneumonia: CT findings of 122 patients and differentiation from influenza pneumonia," European Radiology, p. 1, 2020.

[10] W. Chen, X. Xiong, B. Xie, et al., "Pulmonary invasive fungal disease and bacterial pneumonia: a comparative study with high-resolution CT," American Journal of translational research, vol. 11, no. 7, p. 4542, 2019.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[12] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2018, pp. 7794–7803.

[13] X. Qian, Y. Fu, T. Xiang, Y.-G. Jiang, and X. Xue, "Leader-based multi-scaleattentiondeeparchitectureforpersonre-identification," IEEE Transactions on pattern analysis and machine intelligence, vol. 42, no. 2, pp. 371–385, 2019.

[14] W. Wang, Y. Fu, X. Qian, Y.-G. Jiang, Q. Tian, and X. Xue, "Fm2unet: Face morphological multi-branch network for makeup-invariant face verification," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 5730–5740.

[15] K. Chen, J. Pang, J. Wang et al., "Hybrid task cascade for instance segmentation," in IEEE Conference on computer vision and pattern recognition, 2019, pp. 4974–4983.

[16] https://colab.research.google.com/

[17] https://www.tutorialspoint.com/google_colab/what_is_google_colab.htm

[18] https://www.codingforentrepreneurs.com/courses/python-google-colab-sheets-drive/

https://github.com/pavan-reddy-28/NN-DL_Final_project