Lung disease prediction from X-ray images

Bogdana Zivkovic[†], Nikola Ivanovic[‡]

***Abstract*—The emergence of COVID-19 in December 2019 marked a significant global health crisis. Rapid spread of the virus across the world required urgent diagnostic solutions. Due to multiple challenges that reverse-transcription polymerase chain reaction (RT-PCR) tests face, such as laboratory capacity constraits and costs linked to equipment and PCR reagents, new alternatives for diagnosis had to be found. Chest X-rays present a viable alternative. However, only the experts are able to interpret the abnormalities present in chest X-rays accurately. Due to a shortage of such experts, machine learning models trained on radiography images can significantly aid in identifying these patterns.**

In this paper, we compare various classifying approaches: one built using an encoder derived from a trained autoencoder, Inception V4 and DenseNet-121. Lastly, these methods are merged into an ensemble that boosts overall performance obtaining the accuracy of 95,64%.

***Index Terms*—COVID-19, Pneumonia, Chest X-ray, Deep Learning, Neural Networks, Ensemble, Inception V4, DenseNet-121, Autoencoder.**

I. Introduction

In December 2019, Wuhan City, China, experienced an unprecedented outbreak of pneumonia with an unknown cause. The World Health Organization (WHO) later identified a novel coronavirus as the culprit and named it COVID-19. This virus, belonging to the betacoronavirus family and closely related to severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), primarily affects the lower respiratory tract, leading to pneumonia in humans. Despite extensive global efforts in containment and quarantine, the number of COVID-19 cases continued to escalate. [1]

COVID-19 can be diagnosed with the reverse-transcription polymerase chain reaction (RT-PCR) test. In order to perform this test, it is first necessary to collect suitable respiratory tract samples. Additionally, conducting the RT-PCR test is feasible only in laboratories with the necessary infrastructure. Furthermore, there are instances where the COVID-19 test may need to be repeated within one or two days. This diagnostic procedure is often costly due to the expenses associated with equipment and PCR reagents. RT-PCR test can be time-consuming, requiring specialized microbiologists for test analysis and demanding strict safety measures, such as personal protective equipment, to ensure the well-being of laboratory personnel. [2]

Besides utilizing the RT-PCR test, COVID-19 lung patterns can be detected on standard chest X-rays. This diagnostic procedure is crucial, considering that some areas of the world don't have access to reliable PCR testing. However, due to the subtlety of X-ray visual cues, radiologists encounter significant challenges in accurately detecting and interpreting these subtle indicators. Therefore, there is a pressing need for computer-aided diagnostic systems to assist radiologists in efficiently and accurately interpreting X-ray alterations characteristic of COVID-19. Machine learning models trained on radiography images have the potential to assist radiologists by providing decision support, thereby speeding up the diagnostic process. [3]

This paper introduces three neural network architectures designed to predict COVID-19 and pneumonia in chest X-ray (CXR) images. The study compares these three different classifiers, one of which is constructed using the encoder derived from an autoencoder previously trained on normal, pneumonia, and COVID-19 X-ray images. Additionally, the study evaluates the performance of Inception V4 and DenseNet121 classifiers. Finally, the methods are combined in an ensemble to enhance overall performance.

The structure of the paper is as follows: Section II gives an overview of previous research on COVID-19 prediction using machine learning. Section III describes the processing pipeline. It offers a high-level introduction to the work conducted. Section IV describes the dataset used and explains the preprocessing steps undertaken. Section V provides a more detailed insight into the Learning Framework, while Section VI presents the final results. Finally, in Section VII, concluding remarks are presented, offering insights derived from the study.

II. Related Work

Machine Learning has witnessed considerable growth in recent years. Specifically, in the medical field, it finds applications in various tasks, notably in the classification of cardiovascular diseases and diabetic retinopathy.

Deep learning methods have the ability to uncover image characteristics that may not be initially visible in the original images. In particular, Convolutional Neural Networks (CNNs) have demonstrated to be highly effective in extracting features and learning, leading to their widespread adoption among researchers. Capabilities of the convolutional neural networks (CNN) have enabled medical professionals to utilize them for a multitude of tasks, including diagnosing skin lesions and detecting brain tumors and breast cancer. [4] [5]

There have been various researches demonstrating the effectiveness of applying Deep Learning models on chest X-ray (CXR) images. Some of these researches demonstrated promising results in the diagnosis of pulmonary diseases such as COVID-19 pneumonia. Some of these works include:

[†]bogdana.zivkovic@studenti.unipd.it
[‡]nikola.ivanovic.1@studenti.unipd.it

- CheXNet, a 121-layer convolutional neural network trained on the largest publicly available chest X-ray dataset, ChestX-ray14, for the classification of 14 pulmonary diseases, developed by Rajpurkar et al. [6]
- A technique for automatic detection of Covid-19 pneumonia from CXR images that utilizes pre-trained convolutional neural networks introduced by Muhammad E. H. Chowdhury et al. [5]
- A deep learning model for identifying COVID-19 in CXR images and grouping similar images proposed by Keidar et al. [7]
- A technique for detecting COVID-19 that utilizes various Deep Learning models alongside a support vector machine (SVM) as a classifier. [8]
- COVID-Net, a custom-designed deep convolutional neural network tailored for the identification of COVID-19 cases from chest X-ray images. [9]

III. Processing Pipeline

In our project, we developed a classifier that extracts features using a deep convolutional autoencoder and performs classification on the feature vector utilizing a feed forward neural network. Furthermore, we experimented with segmenting the image into 16 sub-patches and conducted feature extraction and classification on each of these sub-patches individually, followed by aggregating their predictions to derive a final result. The autoencoder and the neural network classifier were trained both separately and together, exhibiting promising results in both cases.

Additionally we evaluated two popular approaches: Inception V4 and DenseNet121 in order to assess their effectiveness compared to to the primary approach and to test the compatibility of the different models by combining their predictions.

Inception architectures are known for their utilization of ”inception modules” in image classification and object detection tasks. Inception modules are blocks of layers that perform parallel processing at different scales and then combine their results. This approach enables the network to simultaneously capture information across multiple levels of abstraction. The Inception V4 architecture represents a convolutional neural network design that advances upon earlier versions within the Inception family. The architecture is characterized by a more streamlined and uniform design, featuring an increased number of inception modules in comparison to Inception V3.

DenseNet is another convolutional neural network (CNN) atchitecture renowned for its exceptional performance in image classification tasks. DenseNet stands for ”Densely Connected Convolutional Networks”. It earns its name by establishing connections between each layer in a feedforward manner. DenseNet distinguishes itself from other CNN architectures through two primary features: dense block structure and transition layers. Dense block structure refers to a specific arrangement of layers within the neural network where each layer is connected to every other layer in a feedforward fashion. Transition layers refer to the layers that are used to reduce the number of parameters and computational complexity within DenseNet architectures.

IV. Data and Features

For this project, we used the COVID19, Pneumonia and Normal Chest X-ray PA Dataset. The dataset is organized into 3 folders (covid, pneumonia , normal). The data was initially split into train, validation and test sets with 3660, 640 and 275 images respectively, stratifying the data by label to ensure a balanced dataset.

Initially, the images were loaded and transformed into grayscale. The images with uneven aspect ratios underwent center cropping, as the regions of interest predominantly resided in the image center. Additionally, all images were resized to the standard dimensions of 224x224 pixels. The only exceptions to this are the approaches that involved splitting the image into subpatches, where the original image was scaled to 256x256 pixels. Following this preprocessing step, normalization was applied to scale the pixel values within the (0,1) range, ensuring compatibility with the neural network methodology. Finally, we determined that performing one-hot encoding on the labels would be the most suitable approach. To enhance performance we employed the TensorFlow Data API, implementing data processing in batches of varying sizes, which were determined by the trained model.

V. Learning Framework

The approaches leveraging a deep convolutional autoencoder for feature extraction require separate training of the autoencoder to accurately reconstruct the image from the code. The encoder comprises four convolutional layers followed by a dense layer, which adjusts the encoder’s output to a predetermined code size (in our case, 64) to optimize the bias-variance trade-off. Conversely, the decoder consists of a dense layer and three convolutional layers that upscale the reconstructed image to its original format. All layers utilize the Exponential Linear Unit (ELU) activation function and employ padding to maintain the desired output size.

The model is configured with a mean squared error loss function and Adamax optimizer, trained for 30 epochs using a batch size of 32. Subsequently, the extracted features are utilized for a feed-forward neural classifier consisting of three fully connected layers with 128, 64, and 3 neurons, respectively. This architecture is chosen for its simplicity and efficient computation, given that the input is a simplified representation of the data.

Two slightly different approaches are employed: the first approach utilizes the previously trained encoder for feature extraction and focuses on training the feed-forward classifier separately, while the second approach trains the encoder again with the classifier.

Methods that rely on feature extraction proved to be quite simple and fast to execute compared with the other models, however their simplicity somewhat limited their accuracy. Once the features were accurately extracted, even very simple classifier models could capture the patterns in the features and

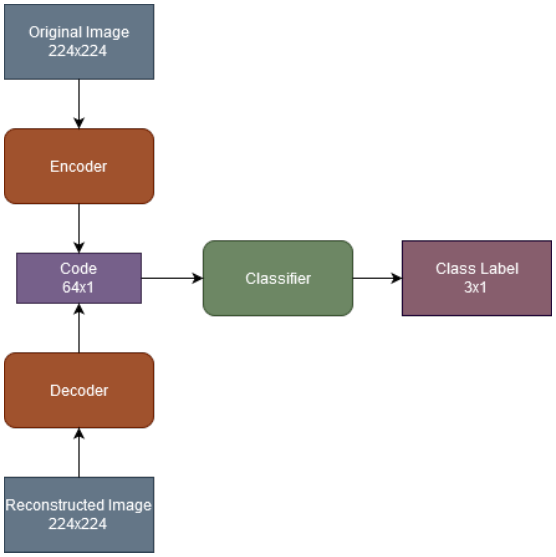

Fig. 1: Training of the Encoder and Classifier

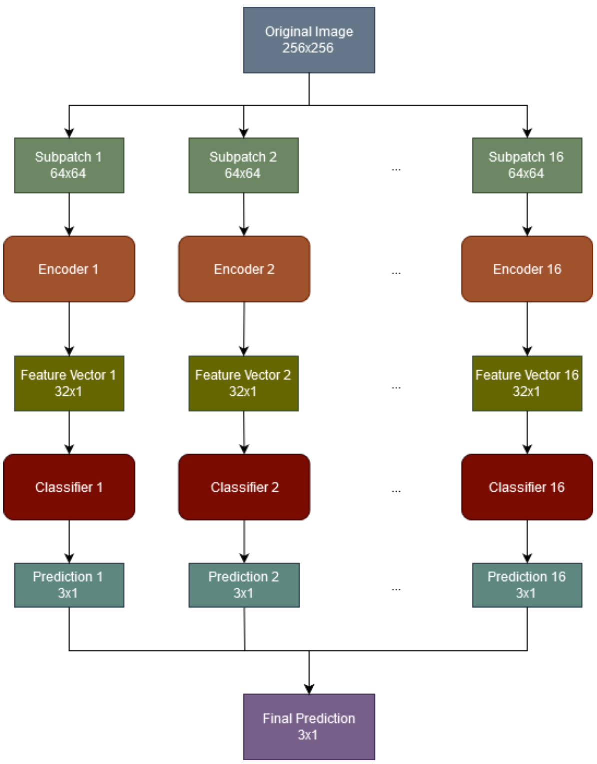

Fig. 2: Classifying Subpatches of the Image

would often even overfit. A straightforward way to expand on this approach in order to arrive at a more complex model was split every image to subpatches and apply the same procedure to each one, aggregating their predictions.

Thus, the original image was scaled to 256x256 pixels and split into sixteen subpatches of 64x64 pixels each. The autoencoder architecture selected for this task was very similar to the one used for encoding the entire image, however it was adapted to be able to reconstruct a smaller image. Subpatches corresponding to the same part of the image were grouped together and a separate autoencoder was trained for every group. In this case, a feature vector of length 32 was employed.

As in the previous approach, the feature vectors were fed to sixteen feed forward classifiers which were trained separately for every group of subpatches. The architecture of the classifiers was also very similar to the general one. As the output logits from the classifiers reflect a degree of certainty for the predicted class, the predictions for the entire image were obtained by simply averaging all of the outputs. Overall, each pair of autoencoder and classifier was simple and fast to train, although fitting all of the models was quite demanding computationally.

The second method evaluated was the Inception V4 model. The architecture can be summarized as follows:

- Input layer
- Stem Block
- Inception-A Block
- Reduction A
- Inception-B Block
- Reduction B
- Inception-C Block
- Global Average Pooling
- Dropout
- Output layer

The Stem Block in this context denotes the initial sequence of operations conducted prior to the introduction of the Inception blocks. Inception blocks A, B, and C in the Inception V4 architecture are specialized components designed to enhance the network's ability to capture features at different scales and complexities within the same layer. Inception V4 introduced dedicated "Reduction Blocks". While earlier versions did not explicitly feature reduction blocks, the same functionality was integrated into their design.

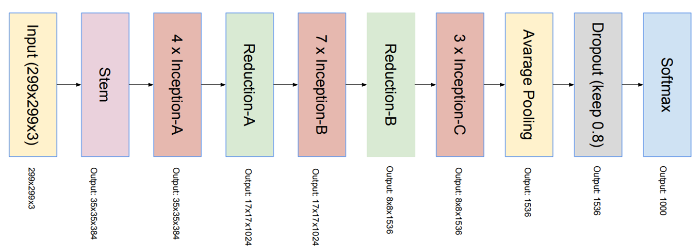

Fig. 3: Inception V4

Next, the effectiveness of DenseNet-121 for this specific problem was assessed. DenseNet-121 is a variant of DenseNet with 121 layers. The architectural layout of DenseNet-121 is outlined as follows:

- Input layer
- Convolution layer with 64 7x7 filters and a stride of 2
- Pooling layer with 3x3 max pooling and a stride of 2
- Dense Block 1 with 2 convolutions repeated 6 times
- Transition layer 1 (1 Conv + 1 AvgPool)
- Dense Block 2 with 2 convolutions repeated 12 times

- Transition layer 2 (1 Conv + 1 AvgPool)
- Dense Block 3 with 2 convolutions repeated 24 times
- Transition layer 3 (1 Conv + 1 AvgPool)
- Dense Block 4 with 2 convolutions repeated 16 times
- Global Average Pooling layer - accepts all the feature maps of the network to perform classification
- Output layer

After the initial input layer, each subsequent convolutional block follows this sequence: Batch Normalization, ReLU activation, and the Conv2D layer.

Each dense block consists of two convolutional operations, utilizing kernels of sizes 1x1 and 3x3. This sequence is iterated multiple times within each dense block: 6 times in dense block 1, 12 times in dense block 2, 24 times in dense block 3, and 16 times in dense block 4.

Within the transition layer, the aim is to halve the number of channels from the current configuration. This involves the incorporation of a 1x1 convolutional layer alongside a 2x2 average pooling layer, with a stride set to 2. [10]

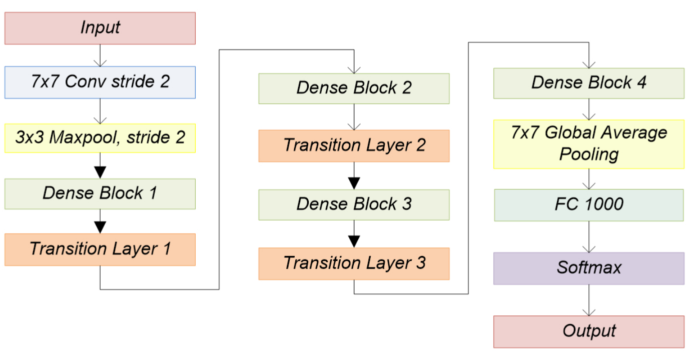

Fig. 4: DenseNet-121

Finally, in order to examine the compatibility of the different approaches, as well as attempt to boost performance, several trained models were integrated in an ensemble. The final predictions of the joined models were simply averaged to obtain the final class.

VI. Results

Although various performance metrics were computed for the examined models, here we will highlight the ones that are the most indicative of the models' performance. Firstly, accuracy assesses the models' general ability to adapt to training data and generalize to unseen data, irrespective of the specific requirements of the problem at hand. Secondly, the recall for the Covid and Pneumonia classes are considered, because these metrics address the usual goal of this kind of system, which is to accurately capture all cases of disease in the lungs, while tolerating faults in the prediction of healthy cases.

Overall, the performance of the models is comparable in all of the chosen metrics. However, many models tended to either predict the cases of Covid-19 or pneumonia worse when compared to other classes, as reflected by the respective recall scores. In this regard, the two known architectures, Inception and Densenet showed better results than most of the experimental approaches. On the other hand, some of the experimental approaches exceed these methods in terms of general accuracy.

Finally, an ensemble of the encoder and classifier model, the Inception model and the Densenet model achieved higher accuracy than any of the models individually, without compromising the ability to accurately detect Covid-19 and pneumonia. This proves that the different approaches are indeed complementary to a certain degree, as their shortcomings are accounted for by the other models.

	Model	**Accuracy**
1.	Encoder + ANN	93.45%
2.	Retrained Encoder + ANN	92.00%
3.	Inception V4	93.82%
4.	Densenet	92.73%
5.	Encoder + ANN on Subpatches	92.36%
6.	Retrained Encoder + ANN on Subpatches	94.55%
7.	Ensemble of Models 2, 3 and 4	95.64%

TABLE 1: Accuracy on the Test Set

	Model	**Recall (Covid)**
1.	Encoder + ANN	95.65%
2.	Retrained Encoder + ANN	88.04%
3.	Inception V4	97.83%
4.	Densenet	95.65%
5.	Encoder + ANN on Subpatches	90.22%
6.	Retrained Encoder + ANN on Subpatches	93.48%
7.	Ensemble of models 2, 3 and 4	96.74%

TABLE 2: Recall (Covid) on the Test Set

	Model	**Recall (Pneumonia)**
1.	Encoder + ANN	89.13%
2.	Retrained Encoder + ANN	93.48%
3.	Inception V4	91.30%
4.	Densenet	91.30%
5.	Encoder + ANN on Subpatches	89.13%
6.	Retrained Encoder + ANN on Subpatches	91.30%
7.	Ensemble of models 2, 3 and 4	91.30%

TABLE 3: Recall (Pneumonia) on the Test Set

	Model	**Time [s]**
1.	Encoder + ANN	650
2.	Retrained Encoder + ANN	750
3.	Inception V4	3500
4.	Densenet	7000
5.	Encoder + ANN on Subpatches	16500
6.	Retrained Encoder + ANN on Subpatches	16500

TABLE 4: Approximate Training Times

In terms of the complexity of the different models, the approaches that involved training the encoder and classifier on the entire image are effective with minimal training. The complexity of Inception and Densenet were comparable, while the training of sixteen separate encoders and classifiers on the subpatches proved to be the most computationally demanding.

It should be pointed out that the more complex cases might achieve better performance with longer training and fine tuning of the parameters, while the simple models are held back by their inherent simplicity.

More details about the implementation can be found on: https://github.com/ivanovicnikola/HDA_final_project

VII. Concluding Remarks

Many modeling approaches proved effective for tackling the problem of diagnosing lung diseases from X-ray images. Feature extraction via deep convolutional autoencoder is particularly effective for image classification as it accomplishes good performance even with minimal training and paired with a simple classifier. Splitting the image into subpatches results in a more robust model, at a much higher computational cost. Overall, classifying subpatches is a promising approach, but requires refining to justify the higher complexity. Both Inception and Densenet achieve comparable levels of accuracy with moderate training procedures and displayed complementary behavior with experimental approaches. Observing that all of the models reached a threshold at around 95% accuracy, we presume that additional preprocessing of the initial dataset is necessary in order to surpass this. However, the goal of this project wasn't to develop a single highly effective model, but rather to experiment with different approaches and analyze their results.

References

[1] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19)," *International journal of surgery (London, England)*, vol. 76, pp. 71–76, 2020.

[2] I. Mporas and P. Naronglerdrit, "Covid-19 identification from chest x-rays," in *2020 International Conference on Biomedical Innovations and Applications (BIA)*, pp. 69–72, 2020.

[3] T. Majeed, R. Rashid, D. Ali, and A. Asaad, "Issues associated with deploying cnn transfer learning to detect COVID-19 from chest X-rays," *Australasian College of Physical Scientists and Engineers in Medicine*, 2020. Received: 17 May 2020 / Accepted: 25 September 2020 / Published online: 6 October 2020.

[4] M. Constantinou, T. Exarchos, A. Vrahatis, and P. Vlamos, "Covid-19 classification on chest x-ray images using deep learning methods," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, p. 2035, 2023.

[5] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can ai help in screening viral and covid-19 pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020.

[6] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv*, 2017.

[7] D. Keidar, D. Yaron, E. Goldstein, Y. Shachar, A. Blass, L. Charbinsky, I. Aharony, L. Lifshitz, D. Lumelsky, Z. Neeman, M. Mizrachi, M. Hajouj, N. Eizenbach, E. Sela, C. Weiss, P. Levin, O. Benjaminov, G. Bachar, S. Tamir, Y. Rapson, D. Suhami, E. Atar, A. Dror, N. Bogot, A. Grubstein, N. Shabshin, Y. Elyada, and Y. Eldar, "COVID-19 classification of X-ray images using deep neural networks," *Eur Radiol*, vol. 31, pp. 9654–9663, Dec 2021. Epub 2021 May 29.

[8] M. Togacar, B. Ergen, and Z. Comert, "COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches," *Comput Biol Med*, vol. 121, p. 103805, Jun 2020. Epub 2020 May 6.

[9] L. Wang, Z. Lin, and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Sci Rep*, vol. 10, p. 19549, 2020.

[10] A. Vidhya, "Unpacking densenet to understand and then creating using tensorflow."