"COVID-19 automated recognition with deep learning approach"

Pavel Ianko†

***Abstract*—The outbreak of COVID-19 disease pushes medical development to propose effective tools for diagnosing and isolating infected patients. A common solution is a polymerase chain reaction (PCR) test, characterised with low sensitivity, laboriousness and time costs. Recently, deep learning models prove to be a novel solution, able to compensate for PCR weaknesses. Deploying a machine learning classifier will reduce workforce demands and allow for automatic, streamlined diagnosing. In this work, five models (CNN, Dropout-CNN, AlexNet, VGG16, VGG19) were trained on X-ray images to discern COVID infection from pneumonia and healthy chest. The best model in this study – AlexNet – together with contrast adaptive histogram equalization (CLAHE), holds classification accuracy of 90%, with 3.5% probability of type II mistake. The rate of type II mistakes is suggested as a critical domain-related metric, which describes the risk to misclassify COVID-infected patients as having other type of lung pathology.**

***Index Terms*—Neural Networks, Deep Learning, CLAHE.**

I. Introduction

Since 2019, the COVID-19 pandemic holds strong detrimental effect on public health and economical development. By the 29th of January, more than five millions of lethal outcomes press society for automating early-stage diagnosis and preventive care.

Today, a polymerase chain reaction (PCR) test is a common method for COVID detection [1]. However, despite the accuracy of the identification by tissue sampling, PCR fails to provide quick result. Hence, here emerges a need for automated disease recognition.

In this paper, we examine deep learning methods for recognising pathological patterns in patients' lungs, based on chest X-ray (CXR) images. Depsite the CXR diagnosis is inaccurate on mild disease stages, the X-ray image production is less time consuming than PCR method [2]. Despite automated recognition models must not be used autonomously for diagnosis [3], they have potential to compensate for low PCR sensitivity (60% - 70%) and dependence on sample purity [4].

This paper aims to raise testing sensitivity with deep learning approach. Deploying a robust COVID classificator, combined with PCR testing, will reduce domain knowledge for healthcare workforce. Also, a preliminary automated sorting of patients' cases is relevant to facilitate quick diagnosis.

This work takes advantage of convolutional neural networks (CNN). As translation-invariant models, they hold capacity for detecting COVID pathologies independent on their location on an X-ray image.

According to the results of this study, deep learning approach is advantageous for compensating PCR low sensitivity and rapid preliminary diagnosis of COVID-19 disease.

- Among five models studied, AlexNet yields classification accuracy, F1, precision and recall metrics of 87%, comparable with previous studies,
- Histogram equalization and increased number of learning epochs increased the AlexNet performance up to 90%,
- In addition to classification metrics, the models' complexity and proneness to type II mistake is examined. The probability of the type II mistake is suggested as a critical factor for selecting the best model for COVID detection.

In the following section we introduce an overview of related works on application of Deep Learning and preprocessing techniques for COVID diagnosis. Section III is dedicated to the scheme of the experiments and followed by models' architecture and data description in chapter IV. The learning procedure and chosen architectures are described in section V. The performance, complexity and domain-related metrics are discussed in the results section (part VI).

II. Related Work

This part reveals measures, taken for automating COVID diagnosis with neural networks approach. A decent part of related work is devoted to transfer learning – adjusting common pre-trained CNN architectures for COVID detection. Thus, Rahaman *et al.* [5] took advantage of diverse architectures (VGG, Xception, ReNet, etc.), pre-learned on Imagenet dataset. Although the authors reported a considerably high accuracy of 89.3% for VGG19 network, it might be improved with contrast enhancing preprocessing techniques, which were not considered, but implemented in this work.

Rather than classifying, Cohen *et al.* [6] focus on the regression task of severity prediction. The authors used DenseNet architecture, with preliminary fit on publicly available CXR images for 18 lung pathologies. In this research, saliency maps demonstrate image critical regions, on which DenseNet focuses for making a prediction. It was shown, how the network pays attention at lung opacities. Although, additional preprocessing, such as histogram equalization, can highlight the opaque regions, raising the regression metrics.

†Data Science Department, University of Padova, email: pavel.ianko@studenti.unipd.it

In addition to conventional models, Wang *et al.* focused on a novel CNNdesign, tailored for COVID detection [7]. The architecture leverages "projection expansion projection extension" (PEPX) blocks, pre-trained on ImageNet dataset. The model was re-adjusted on COVIDx dataset of over 13,000 CXR samples. With data augmentation techniques (crop, image flip, zoom and shift), the model reaches 93.3% test accuracy and higher sensitivity, as opposed to VGG-19 and ResNet-50.

A share of related works is devoted to contrast enhancement and preprocessing, as noise and light scarcity in CXR images undermine models performance. In the work of Santra [8], utilizing white balance and adaptive histogram equalization yields 96.43% accuracy for a Depth-Wise Separable (DSCNN) network. Trained on a relatively small set of 1,823 CXR samples, DSCNN outperforms the majority of models for COVID recognition (VGG19 [9], COVID-Net [7], concatenation of Xception and ResNet50V2 [10]). In the studies of Narin *et al.* [11], the problem is approached as a binary classification with state-of-the-art models (InceptionV3, ResNet-50, 101, 152, V2). The offered solution holds over 99% accuracy for discrimination between COVID and viral pneumonia, COVID and bacterial pneumonia. However, the validity of using accuracy is debatable, as the authors use imbalanced datasets.

According to the related research, the authors are either focused on state-of-the art architectures, or take advantage of contrast enhancing techniques. However, combining both approaches has potential for increasing sensitivity of the network, and thus is considered in this paper.

III. Processing Pipeline

The goal of this paper is to combine deep learning architectures with advanced preprocessing techniques, for raising the classifiers' sensitivity for COVID-19 recognition. The conducted experiments aim to compare models' performance after applying CLAHE histogram equalization method. Thus, figure 1 illustrates a general experiment scheme. Since the evaluation metrics are credible, as long as the data is balanced, we start with a balanced split of available data into train, validation and test sets.

Further, the experiment is divided into two stages. The first stage assesses the performance of several CNN-based models. To allow for greater generalisation ability and avoid overfitting, we take advantage of data augmentation and simple scaling, yet not using contrast enhancement. Several performance metrics (accuracy, F1, precision and recall) are used to select robust architectures. Finally, the histogram equalization technique is applied to the best model, to evaluate the improvement after contrast enhancement [8].

Moreover, the domain of the problem suggests additional metrics to consider. One has to account for model complexity, learning and response time. The more a model is lightweight, the higher its capacity for rapid response and readjustment with more training data. Second, the least desirable mistake is a type II error, when a patient is classified as non-COVID, while suffering from this disease. This is a crucial metric to minimize, and is used to select the best model.

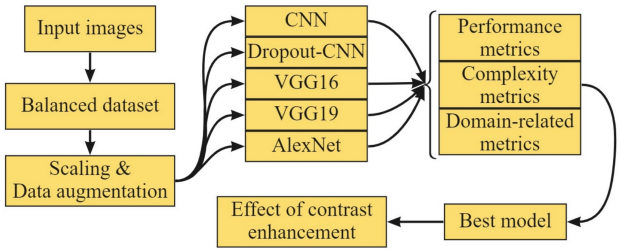

Fig. 1: High level scheme of the experiment

IV. Signals and Features

The available dataset is a collection of 4575 images, divided into three categories of patients – infected with COVID, pneumonia, or having normal chest. The data is randomly split into train, validation and test sets (3201, 687 and 687 images respectively), maintaining 70%, 15% and 15% proportion (Fig. 2).

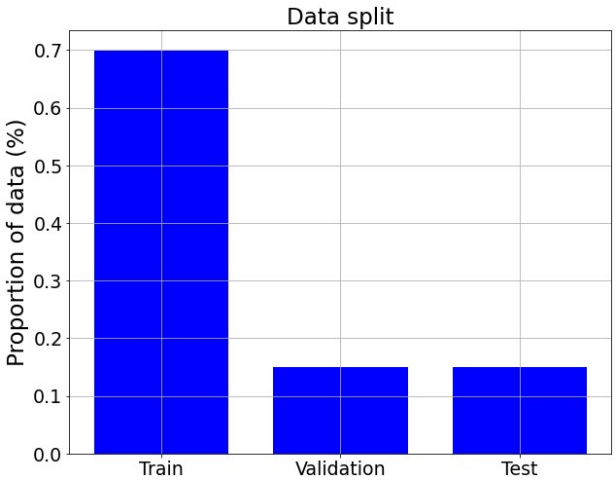

Fig. 2: Split between train, validation and test data (3201, 687 and 687 images respectively)

Instances of each of the three classes are uniformly distributed within each dataset (Fig. 3), for credibility of the proposed evaluation metrics.

Figure 4 demonstrates examples of images, corresponding to each class. From the images it is clear how contrast enhancing methods highlight pathological structures in lungs. However, the difference between pneumonia and COVID is barely visible, requiring domain knowledge and expertise to diagnose a patient.

V. Learning Framework

A. Architectures

In this study, several models are considered:

- Simple CNN,
- CNN with dropout,

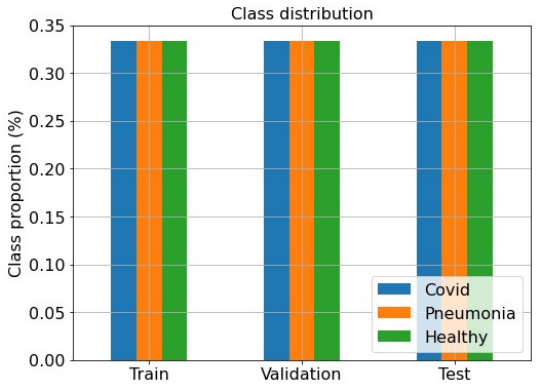

Fig. 3: Distribution of classes within train, validation and test sets. A uniform distribution is maintained, for credibility of evaluation metrics (accuracy, sensitivity, etc.)

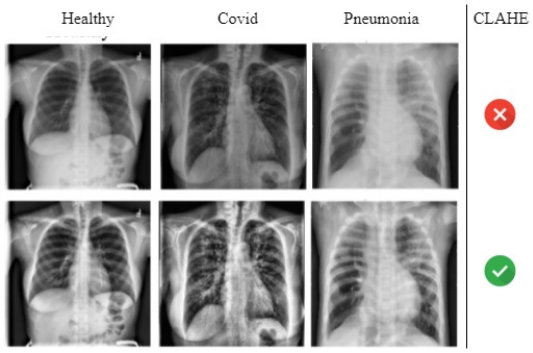

Fig. 4: Examples of train images, corresponding to three types of patients (having a normal chest, either ill with COVID or pneumonia)

- AlexNet,
- VGG16,
- VGG19.

We begin with the simplest CNN model, taking advantage of translation-invariant convolutional blocks, followed by activation and pooling layers (Fig. 5). The issue of zero gradients during back propagation is avoided by using leaky ReLU activation function. Pooling layers help to reduce dimensionality of feature maps, selecting the most intense visual concepts from convolutional blocks.

Kernel size was fixed to 3×3 for all layers, completed with HeNormal initialization. L1 and L2 regularization terms (0.01 each) avoid overfitting problem.

Guided by Park *et al.* [12] and Hinton *et al.* [13], kernel regularization was substituted with dropout layers in Dropout-CNN model (Fig. 6). While convolutional blocks are followed a slight dropout rate of 10%, a 40% of neurons is randomly discarded at the output layer.

According to related works, high performance metrics were reported for transfer-learned commonly known architectures (AlexNet, VGG16, VGG19, ResNet, Inception, etc. [5], [8], [14]). Thus, a range of state-of-the-art models was deployed in this work. AlexNet (Fig. 7) is used with HeNormal weights initialization and ReLU activation function.

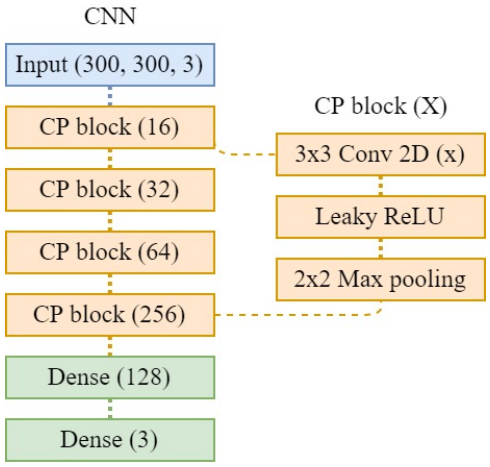

Fig. 5: Architecture of CNN model

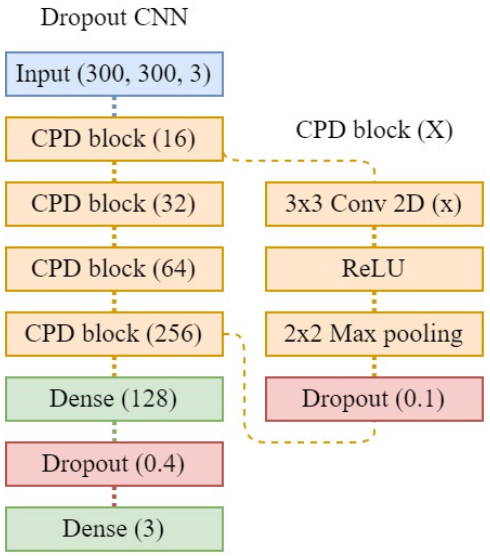

Fig. 6: Architecture of a Dropout CNN model

VGG16 (Fig. 8) and VGG19 (Fig. 9) models are constructed with the same weights initialization and activation functions, as AlexNet.

B. Learning strategy

From visiual inspection of the data, one can conclude that position of the patients' lungs shifts randomly from image to image. Moreover, a part of the data is rotated and flipped upside down. This is why, to raise the generalisation ability and avoid overfitting, we apply data augmentation techniques to the train set.

As suggested by Rahaman *et al.* [5], during training a random horizontal shift (9% of width), vertical shift (9% of height), horizontal and vertical flips are enabled. In addition, we impose random rotation range of 20% and a zoom range of $\pm 10\%$. Thus, we ensure that each epoch a model is evaluated on new batches of modified images. No augmentation is applied to validation and test sets.

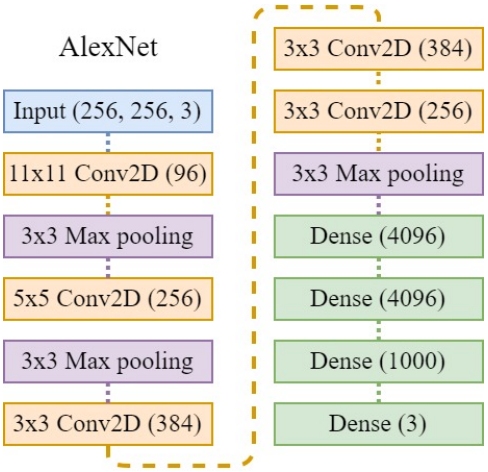

Fig. 7: Architecture of an AlexNet model

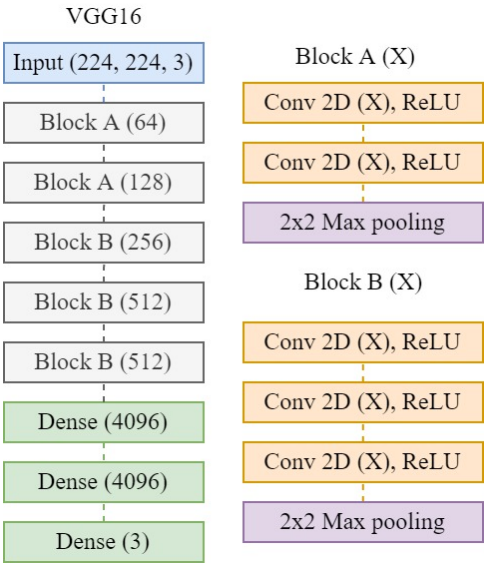

Fig. 8: Architecture of a VGG16 model

Learning algorithm is a stochastic gradient descend, with shuffled train set, divided into mini-batches of 32 images. Validation and test images are not shuffled, since no weight update takes place after evaluation on these sets.

During learning, early stopping mechanism is used to save temporal and computational resources. At the end of each epoch, early stopping callback monitored categorical crossentropy loss on the validation dataset. If the loss did not decrease during two epochs in a row more than 0.0007, the learning was stopped, with saving the last epoch weights.

Optimizers and learning rates were selected experimentally for each model, as they greatly impacted the learning process. Thus, AlexNet and VGG16 worked with SGD optimizer and learning rates 0.001 and 0.00001 respectively. Conversely, Dropout and Dropout-CNN models performed well with Adam optimizer and 0.001 learning rate.

Final models comparison is conducted on a test set, with

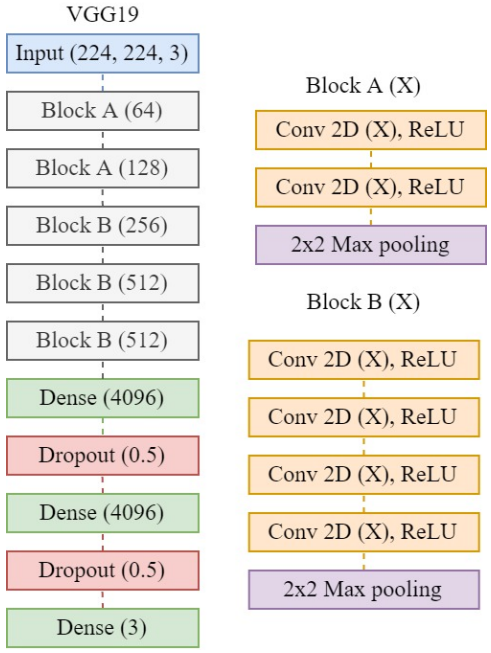

Fig. 9: Architecture of a VGG19 model

no data augmentation and shuffling applied. The evaluation metrics, given in the order of their importance, are:

- Type II mistake risk,
- Recall (Sensitivity),
- Accuracy,
- Precision,
- F1,
- Trainable parameters,
- Evaluation time,
- Time per epoch.

Finally, we follow the same learning scheme, yet with applied CLAHE method for contrast enhancement. This will clarify how advanced preprocessing techniques impact the models performance and inhibit learning and evaluation time.

VI. Results

In this section, we start with reporting learning curves, to reassure the validity of chosen regularization and fitting techniques. To evaluate both performance and models complexity, chosen metrics are divided into two parts and reported for all models. This is followed by examination of the type II mistake risk. The section is finalized with the effect of histogram equalization and interpretability of CNN-based models.

A. Learning summary

For each of the studied models, figures 10 and 11 summarize learning curves on train and validation sets. As there is no remarkable discrepancy between train and validation metrics, no overfitting was observed. However, the roughness of the learning curves suggests that decreasing

learning rate for VGG16, Dropout and AlexNet models could improve the learning quality.

In addition, the steepness of VGG19 (fig. 11) and AlexNet (fig. 10) curves indicates the potential for achieving higher accuracy, if less restricting early stop conditions were applied.

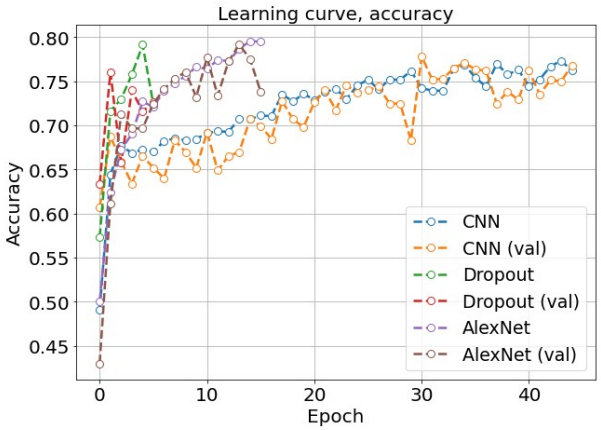

Fig. 10: Learning curves summary for CNN, Dropout-CNN and AlexNet models. "Val" stands for accuracy on the validation set during training

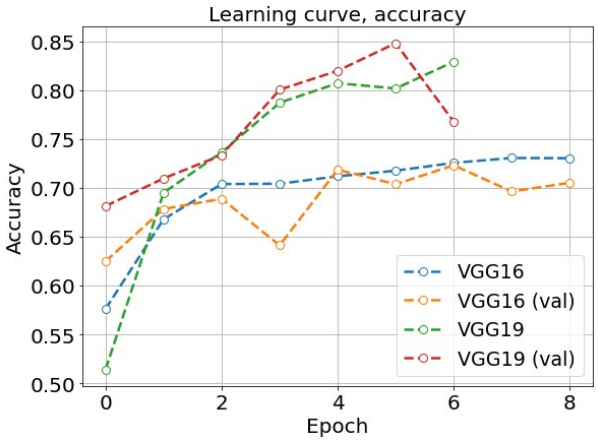

Fig. 11: Learning curves summary for VGG16, VGG19 models. "Val" stands for accuracy on the validation set during training

Bar chart on the figure 12 shows, that setting less restrictions on learning process (e.g., no monitoring of minimal delta of validation loss) results in much higher number of epochs. However, kernel regularization prevented CNN model from overfitting.

B. Complexity metrics

1) Models complexity: Figure 13 summarizes the number of trainable parameter, crossentropy loss on the test set, and the time, necessary for computing test predictions. Because of different ranges, each metric was max-normalized.

The complexity metrics demonstrate a remarkable growth of complexity with the depth-wise and width-wise model growth.

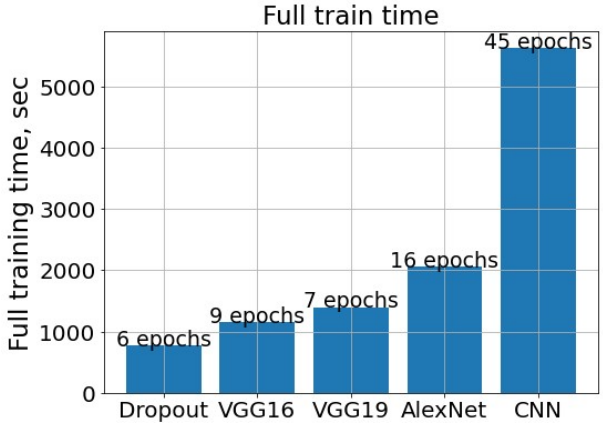

Fig. 12: Summary of train time and number of epochs for the studied models

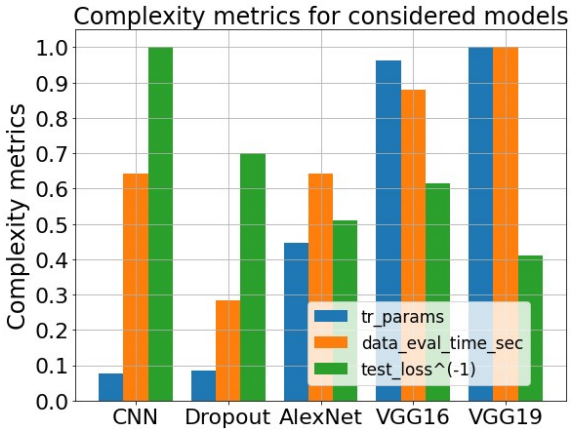

Fig. 13: Max-normalized complexity metrics, evaluated on the test set. "Tr˙params" stands for number of trainable parameters, "data˙eval˙time˙sec" – number of seconds, necessary for computing test predictions, "test˙loss^(-1)" - inverse loss on the test set

The most lightweight model (CNN) stands for less than 10% of VGG19 complexity. Conversely, CNN model achieved the least loss on the test set, as opposed to VGG19. Based on these metrics, AlexNet model has compromised traits, with the medium complexity, test loss and evaluation time.

2) Time requirements: Figure 14 demonstrates a noticeable gap between time, necessary for models to accomplish a learning epoch. This correlates with previous results on Fig. 13, as the most complex model – VGG19 – takes approximately 1.5 times more temporal resources to finish an epoch.

C. Performance metrics

Accuracy, F1, precision and recall are chosen as valid metrics for a balanced dataset. A special attention is drawn towards recall, as this metrics characterizes a model's robustness to type II mistake and ability to patients. As shown on figure 15, all considered models give consistent results.

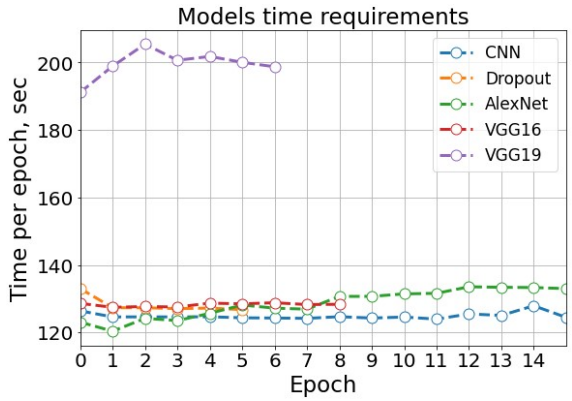

Fig. 14: Number of seconds, necessary for models to accomplish one epoch

VGG19 and AlexNet stand out with the highest metrics slightly above 87%. However, according to the complexity diagrams (Fig. 13, Fig. 14), AlexNet is almost 40% simpler than VGG19, in terms of trainable parameters. This result suggests AlexNet as a proper model for testing with contrast enhancement.

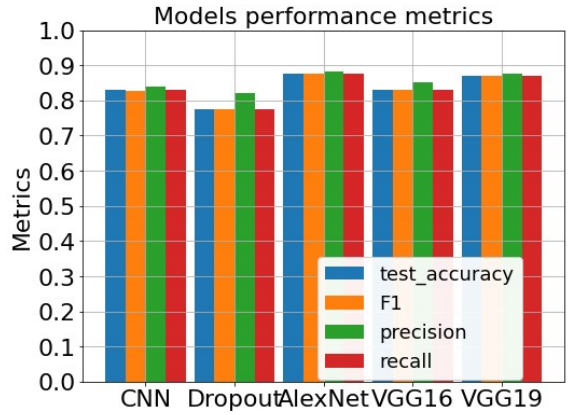

Fig. 15: Accuracy, F1, precision and recall of the studied models

D. Domain-related metric

The crucial trait of an auto-recognition model for medical applications is the ability for capturing infected patients. In case of such an contagious disease as COVID-19, diagnosing a healthy person as ill is not as critical, as labelling a COVID-infected healthy. This misdiagnosis is known as a type II mistake, and considered in this chapter.

An example of confusion matrix, calculated on the test set, is presented on figure 16. The right upper square describes the proportion of classes, labeled "Healthy", while relating to COVID-19 class. This is interpreted as a probability of the type II mistake.

For all the studied models, confusion matrices are calculated on the unseen test data (Tab. 1). Despite of the acceptable test performance, CNN and VGG16 models (Fig. 15), fall out of consideration, for over 25% of misclassifications. Despite the

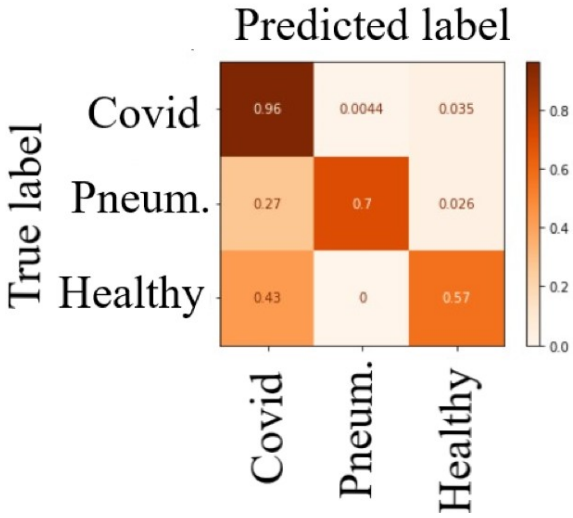

Fig. 16: Confusion matrix for AlexNet model. The right upper square signifies the type II mistake probability of 3.5%

Dropout model has low type II mistake rate, it underperforms on the test set (Fig. 15). AlexNet and VGG19 exhibit the lowest probability of such mistake, despite the great difference in complexities (Fig. 13).

Model	Type II test mistakes, %
AlexNet	**3.5**
CNN	25
VGG16	31
VGG19	**6.2**
Dropout	**5.7**

TABLE 1: Probability of type II mistake for studied models

Based on previous observations, AlexNet model achieves 87% accuracy and recall, with the least proportion of type II mistakes. Meanwhile, this model has acceptable complexity (Fig. 13) and training time requirements (Fig. 14). Hence, we choose this model to study the effect of adaptive histogram equalization.

E. Effect of contrast enhancement

Contrast limited adaptive histogram equalization (CLAHE) is a preprocessing method for intensifying image contrast. The method has already proved potential for raising the models accuracy in COVID detection [8], as it highlights pathological structures in lungs (Fig. 4).

To test the CLAHE effect, we trained AlexNet of the same configuration, while histogram equalization was applied after max-scaling, during the preprocessing step.

Table 2 compares performance of AlexNet before and after CLAHE application. For now, contribution of CLAHE remains debatable. In a given learning configuration,

CLAHE is proved to raise overall performance by 3%, although increased the probability of type II mistake.

Metric	No CLAHE	CLAHE
Accuracy, %	87.6	90
Type II mistakes, %	3.5	**12**
Recall, %	87.6	**90**
F1, %	87.7	90
Precision, %	88.1	90.2
Epochs	16	21

TABLE 2: Probability of type II mistake for studied models

Compared to previous AlexNet confusion matrix (Fig. 16), CLAHE and increased number of epochs reduced overall models' mistakes rate (Fig. 17).

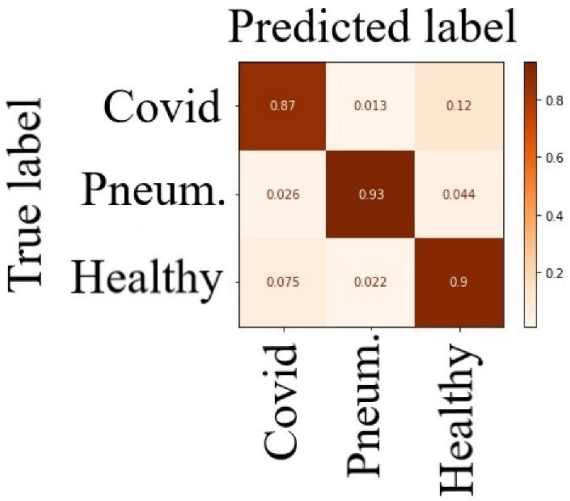

Fig. 17: Confusion matrix for AlexNet model, after applying CLAHE as a preprocessing step

However, the contribution of preprocessing remains debatable. As indicated on figure 19, larger number of learning epochs had greater impact on the model's performance.

Furthermore, applying CLAHE method increased the learning time twice. This suggests, that the use of advanced preprocessing techniques shall be supported by remarkable performance gain. Otherwise, only scaling can save computational time and provide competitive results.

F. Kernel visualization

This part is devoted to interpretability of CNN-based models. Deep learning models are deemed as "black box" systems, as it is unpredictable, when and why a model will make a mistake. This is why the question of deep learning, applied to medicine, remains debatable. However, such visualizations, as on figure 20, allow to understand the regions where a CNN-based model focuses to make a prediction.

Moreover, the translation invariant convolution operation enables the algorithms to recognize pathological patterns anywhere on the CXR image. One can see, how, with the deeper network levels, opaque regions are highlighted and used as evidence for a diagnosis.

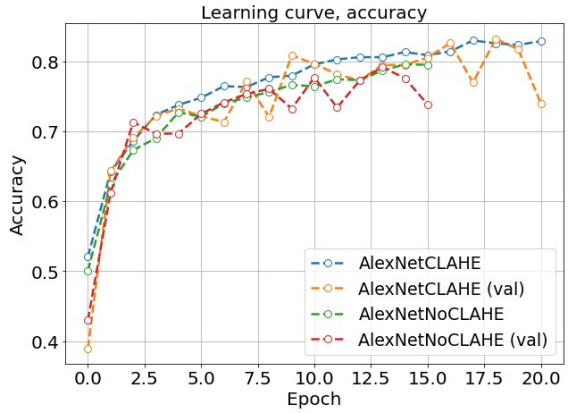

Fig. 18: Comparison of learning curves for AlexNet model, before and after applying CLAHE contrast enhancement

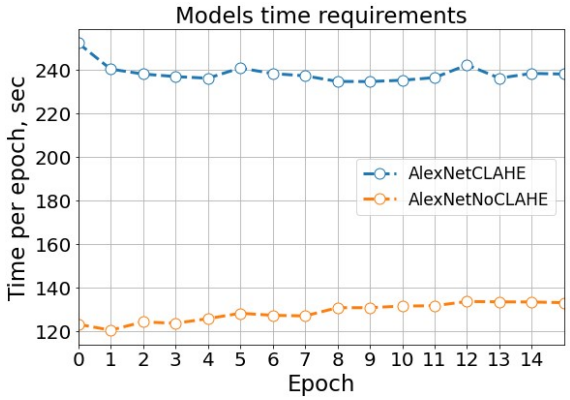

Fig. 19: Comparison of learning time per epoch for AlexNet model, before and after applying contrast enhancement

G. Comparison with the state of the art

Table 3 summarizes performance metrics, corresponding to the previous solutions and the best model of this study. It is seen from the table, AlexNet with CLAHE preprocessing is comparable with VGG16, VGG19 and ResNet. As opposed to that, the proposed model yields to complex architectures, pre-tuned with transfer learning approach [5], [15].

In addition, related studies do not consider a type II probability as a critical parameter to select a model. Nonetheless, this paper suggests that using this metrics helps to discriminate between models and choose the one with least probability of classifying COVID-infected people as healthy or ill with another type of lung disease.

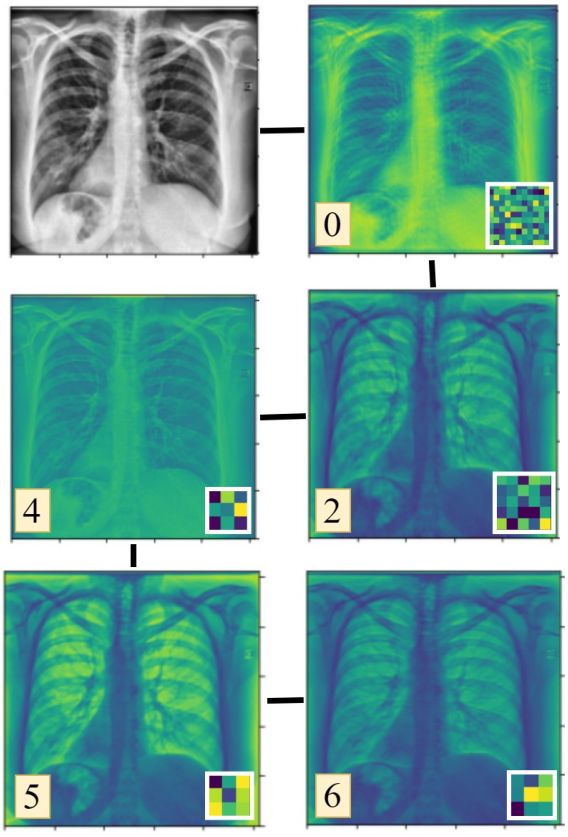

Fig. 20: Visualization of convolutions with kernels at different levels of depth for AlexNet model (green color channel). CLAHE preprocessing is applied to the CXR image of COVID-infected patient. Numbers on the left lower corner indicate the layer, while the right lower corner visualizes the kernel weights

Model	Accuracy	F1	Precision	Recall
DSCNN [8]	96.43	96.00	97.00	96.00
VGG16 [5]	88.57	88.6	88.9	88.6
VGG19 [5]	89.30	89.60	90.80	89.60
ResNet 101 V2 [5]	84.28	84.3	84.7	84.6
Xception + Deep CNN [15]	99.5	98.6	99.4	99.1
Bayes SqueezeNet [16]	98.26	98.25	98.26	98.26
AlexNet + CLAHE	**0.90**	**0.90**	**0.92**	**0.90**

TABLE 3: Comparison with the state-of-the-art models

VII. Concluding Remarks

1) Conclusions: In this study, five CNN-based architectures (CNN, Dropout-CNN, VGG16, VGG19, AlexNet) were applied to a multiclassification task, to discriminate between healthy, COVID-infected and pneumonia-infected patients. 4575 CXR images were split into three balanced sets for training, validating and testing the proposed algorithms.

Selected models were compared, based on four performance metrics with macro averaging (accuracy, F1, precision and recall). In addition, the number of trainable parameters, full training time and seconds per epoch provided insight into models complexity. The study shows that domain-related metrics – probability of type II mistake – is a critical parameter for selecting the best model, robust to misclassifying a COVID-infected patient as healthy. However, to the best of the authors' knowledge, this mistake rate is overlooked in the previous studies, and deserves detailed attention.

AlexNet model was selected to test the effect of contrast enhancement methods on the model's performance. Thus, contrast limited histogram equalization (CLAHE), together with larger number of training epochs, increased the performance metrics (accuracy, F1, precision, recall) up to 90%.

In comparison with the state-of-the-art models, AlexNet with CLAHE enhancement yields comparable result, however fails to compete large-scale transfer-learned models. Based on previous work and conclusions of this paper, a missing remedy to COVID classification is a complex solution, embracing:

- Large-scale deep models (Xception, DSCNN, etc.),
- Transfer-learning with publicly available CXR data,
- Advanced preprocessing techniques (white balance, histogram equalization),
- Using type II mistake as discriminative metrics.

2) What was learned: This project enabled author to learn:

- Common deep learning architectures,
- Saving models and checkpoints,
- Practice building learning curves and data augmentation on practice,
- Working with tensorflow API,
- Using github to coordinate multiple notebooks,
- Enabling object-oriented approach, to keep notebooks shorter,
- Enabling regularization techniques,
- Tracking metrics and time resources,
- Paying attention to diverse metrics.

Difficulties encountered:

- Organizing data flow from google drive directories,
- Finding proper optimizers and learning parameters,
- Using large scale models (ResNet, DenseNet and Inception V5 still refuse to learn from the data).

REFERENCES

[1] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, *et al.*, "Detection of 2019 novel coronavirus (2019-ncov) by real-time rt-pcr," *Eurosurveillance*, vol. 25, no. 3, p. 2000045, 2020.

[2] M. Teymouri, S. Mollazadeh, H. Mortazavi, Z. N. Ghale-Noie, V. Keyvani, F. Aghababaei, M. R. Hamblin, G. Abbaszadeh-Goudarzi, H. Pourghadamyari, S. M. R. Hashemian, *et al.*, "Recent advances and challenges of rt-pcr tests for the diagnosis of covid-19," *Pathology-Research and Practice*, p. 153443, 2021.

[3] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.

[4] M. J. Binnicker, "Challenges and controversies to testing for covid-19," *Journal of clinical microbiology*, vol. 58, no. 11, pp. e01695–20, 2020.

[5] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, M. A. Rahman, Q. Wang, S. Qi, F. Kong, X. Zhu, and X. Zhao, "Identification of covid-19 samples from chest x-ray images using deep learning: A comparison of transfer learning approaches," *Journal of X-ray Science and Technology*, no. Preprint, pp. 1–19, 2020.

[6] J. P. Cohen, L. Dao, K. Roth, P. Morrison, Y. Bengio, A. F. Abbasi, B. Shen, H. K. Mahsa, M. Ghassemi, H. Li, *et al.*, "Predicting covid-19 pneumonia severity on chest x-ray with deep learning," *Cureus*, vol. 12, no. 7, 2020.

[7] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[8] M. Siddhartha and A. Santra, "Covidlite: A depth-wise separable deep neural network with white balance and clahe for detection of covid-19," *arXiv preprint arXiv:2006.13873*, 2020.

[9] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.

[10] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2," *Informatics in Medicine Unlocked*, vol. 19, p. 100360, 2020.

[11] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications*, pp. 1–14, 2021.

[12] S. Park and N. Kwak, "Analysis on the dropout effect in convolutional neural networks," in *Asian conference on computer vision*, pp. 189–204, Springer, 2016.

[13] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[14] E. Cortés and S. Sánchez, "Deep learning transfer with alexnet for chest x-ray covid-19 recognition," *IEEE Latin America Transactions*, vol. 19, no. 6, pp. 944–951, 2021.

[15] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays," *Irbm*, 2020.

[16] F. Ucar and D. Korkmaz, "Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images," *Medical hypotheses*, vol. 140, p. 109761, 2020.