Deep Learning Image Classification for Pneumonia Detection

Teodor Boyadzhiev
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria
t.boyadzhiev@math.bas.bg

Simeon Tsvetanov
Faculty of Mathematics and Informatics
University of Sofia
Sofia, Bulgaria
tsvetanov@fmi.uni-sofia.bg

Stela Dimitrova
Faculty of Mathematics and Informatics
University of Sofia
Sofia, Bulgaria
stelasd@uni-sofia.bg

***Abstract*— Classification of medical images has an important role in medical diagnosis support. Recently deep learning methods have achieved great success at classification of medical images. This paper shows the methodology and experimental results of applying deep learning methods for pneumonia detection from X-rays images. The available dataset was classified using a convolutional neural network. The encoder part of the network follows UNet architecture, which is followed by two fully connected layers.**

The proposed architecture achieved median F1-Score on the testing set of 0.938 and median accuracy of 0.965.

Keywords— deep learning, image classification, pneumonia detection, decision support systems

I. INTRODUCTION

Machine learning can be used to extract structured information from unstructured medical data such as raw text, images and patient records [1].

When the structured data is extracted it can be used for predictive analysis where techniques such as clustering, classification and non-linear regression can be employed.

The first step of a machine learning workflow is data collection and cleaning. If the algorithm is supervised it would also require data labelling. In this step it is important to collect sufficient quantities of data to ensure generalization [2].

The second step is feature extraction. This could happen by manually designing domain specific features or by using non-supervised techniques for automated feature extraction. In the next step a model is designed, trained and tested. Next the model has to be evaluated by a separate dataset, which is not used in the previous step. The evaluation is done according to different metrics, which depend on the use case. Finally if the model meets the quality requirements it is integrated into the clinic workflow [3].

The main contribution of artificial intelligence method is in predictive analysis. Some of the use cases are: evaluation of health risks and incidents, estimation of expected readmission period, evaluation of treatment course and costs. Also identification of high-risk surgical patients, patient's eligibility for clinical trials, symptom significance and help with patient scheduling or predicting no-shows.

Deep learning techniques are widely used for image classification allowing to build advanced decision support systems.

Pneumonia is common disease affecting the lungs and is caused by bacteria or virus infection. Some people who catch COVID-19 get severe pneumonia in both lungs that can be deadly.

The timely and accurate diagnostic is a critical factor for preventing a serious illness, so one use case where machine learning methods can be used is in diagnosis support. For example a medical image can be processed by a neural network to identify anomalies, and different types of structures. Then these areas can be highlighted which will help the medical professional to make the right decisions.

Ayan et al. [4] approached the the xray classification problem by utilising transfer learning. They used the pre-trained networks VGG16 and Xception to classify pneumonia, achieving 87% accuracy.

Hashmi used weighted approach to combine the predictions from ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3. The networks were trained using transfer learning. The method achieved test accuracy of 98.43% on the data from the Guangzhou Women and Children's Medical Center pneumonia dataset [5]

Sirazitdinov et al. used ensamble of two convolutional neural networks, RetineNet and Mask R-CNN to achieve 0.775 f1 score on 26,684 images from Kaggle Pneumonia Detection Challenge [6].

Rahman et al. achieved 98% accuracy on 5247 x-ray chest images by using transfer learning from various networks such as AlexNet, ResNet18, DenseNet201, and SqueezeNet [7]. On the same dataset Hamoldi et al. tried different transfer learning aproaches and custom tailored neural network and achieved 95.72% accuracy [8].

The goal of this work is to determine automatically if a patient has pneumonia based on computer tomography (CT) scans or x-ray images. The problems of automatic diagnosis is approached as a classification problem, for which a convolutional neural network (CNN) is used.

The rest of this paper is organised as follows. Section II provides details about the dataset, network architecture, augmentation and training parameters. Section III describes the results and section IV is the discussion and conclusion.

II. METHODOLOGY

The dataset contains x-ray images of lungs. There are 1583 images taken from different non- pathological patients and

978-1-6654-9578-3/22/$31.00 ©2022 IEEE

4273 images from 1674 patients diagnosed with pneumonia [9].

The methodology is affected by the specifics of the data:

1. There are occasions where several images are taken from the same patient. This violates the assumptions for independently and identically distributed data points, and therefore effectively reduces the dataset size.
2. The patients in the training and testing data set are different, however are given overlapping ids during the anonymization process.
3. It is unclear whether every image in the non-pathological category (NORMAL) is from different patient.
4. The images which pneumonia are labelled as bacterial and viral.
5. The images are compressed with the JPEG algorithm, which causes loss of information and artefacts.
6. The images have different resolutions.
7. The images are at different scales. In some images is visible smaller portion of the chest as compared to other images.
8. The images are centred differently, and patients have slightly different gestures. Some patients have their arms up, other have them to the side.
9. There are writings on the images, such as the letter "R" at different locations. Sometimes there is date or other information.
10. There are foreign objects in the images. Whether they are implants or not is currently unclear.

The data has relevantly low variance, which occurs from:

- The natural variance from the human anatomy
- Insertion of foreign objects (Point 10)
- Writings on the images (Point 9)
- Different poses (Point 8)
- Different imaging techniques and sources (Point 7, and 6)

Therefore, augmentation is used:

- Random rotation from -45 to +45 degrees
- Random zooming
- Random horizontal flip.

Each network was trained 5 times, and the dataset is split into training and testing set before each trail. This way there can be estimated some confidence intervals of the results and can be determined how much of the results is a random noise due to the random split of the data.

The dataset was re-scaled to resolution of 256 by 256 pixels with one grayscale channel and classified by 4 neural networks which use an encoder with architecture of UNet [11] .

Each network has 5 levels, where each level, l, has two 3x3 convolutions, followed by a 2x2 max-pool layer. The activation function for all convolutions and the first fully connected layer is ReLU, since it works better for deep networks than sigmoid [10]. The activation for the second fully connected layer is softmax.

After the last level the resolution is 8x8. The result of the last max-pool layer is reshaped and followed by two fully connected layers. The first fully connected layer has c units with ReLU activation, and the last fully connected layer has softmax activation and has two units, one for normal category and one for pneumonia category.

The number of filters , N(l), in the convolutions in each level is determined by the formula.

$$N(l) = \left\lfloor f 2^{\frac{l}{d}} \right\rfloor \quad (1)$$

where l is the number of level (the count starts from 0), d is parameter called divider, and f is the number of filters in the level 0.

The 4 networks follow the same architecture, but have different meta-parameters:

- tiny - f = 8, d = 4, c = 100
- small - f = 8, d = 3, c = 250.
- medium - f = 8, d = 2, c = 500.
- large - f = 16, d = 2, c = 1000.

where c is the number of units in the fully connected layer.

All networks are initialized with the Xavier [13] method, where parameters are drawn from Gaussian distribution. Each network is trained with Adamax [12] for 50 iterations.

The dataset is split into training and testing set with 70% in the training set and 30% in the testing set. The split is done randomly, by making sure that no images from the same patient will end up in both training and testing set.

The experiment is repeated 5 times to find out how much variance is caused by the splitting of the data-set and over-fitting. The error function was the cross-entropy.

All of the networks used spatial dropout. For the levels 0, and 1 no dropout was used. For level 2 a dropout of 0.1, level 3 a dropout of 0.15, level 4 a dropout 0.2 is used. For the first dense layer the dropout rate is 0.25.

The dropout for the deeper levels, with lower resolution, was higher than the dropout for the shallower levels, higher resolution, because the deeper levels have more filters and therefore have higher change to over-fit the data. The exact numbers were chosen through the process of trial and error.

III. Experiments

Fig 1 Distribution of training and testing F1 score for each network.Shows the distribution of the training and testing F1-Score for each network. The F1-Score is chosen from the epoch with best performance on the testing set, rather than the last epoch.

The results show that the median F1-Score for all of the

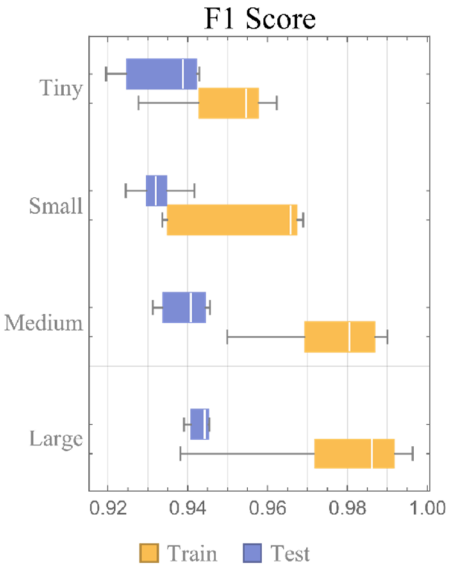

Fig 1 Distribution of training and testing F1 score for each network.

networks is around 0.94, however the discrepancy between the training and testing performance shows great overfitting for the networks "Large" and "Medium". The networks "Small" over-fitted the data to less extend, however the network "Tiny" has better performance than "Small" on the testing set, possibly due to better generalization.

The network "Tiny" was trained on "NVidia GTX 3060" for around 5 minutes. The network makes around 86.6M multiplications when processing an image and has 115 407 parameters.

Network	Accuracy	Precision	Recall	F1
Tiny	96.5	94	93.6	93.8
Ayan	87			
Hashmi	98.43	98.26	99.00	98.6
Sirazitdinov	0.838	75.8	79.3	77.5
Rahman	98	97.0	99.0	97.9
Hamoldi	95.72			

IV. Conclusion

The results in this paper showed that diagnosis of pneumonia from X-Ray images of lungs can be achieved reliably. The use of techniques similar to bootstrapping helped to find confidence intervals for the performance measures of the classifier, making the conclusions more reliable.

Although the median accuracy reported in this paper, 96.5%, is lower than Rahman, 98%, and Hashmi, 98.43%, the number of parameters of "Tiny" is much smaller, 115 407, than the parameters of networks such as VGG16, ResNet18, Xception, InceptionV3, DenseNet121, and MobileNetV3. Thus providing much more computationally efficient solution to the problem with small loss of accuracy, 2%.

The classification suffers from limitations, due to the small dataset used in this study. These limitations have been partially overcome by using dropout and augmentation.

Further research could benefit from larger datasets, which would allow to create more accurate and reliable classifier than the presented in this paper, average accuracy 96.43% and standard deviation of 0.28%.

Acknowledgment

This paper is partially supported by UNITe project: https://unite-bg.eu

References

[1] W. Jialun, T. Kaiwen and Z. Haichuan, "Structured Information Extraction of Pathology Reports with Attention-based Graph Convolutional Network," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 16-19 Dec. 2020.

[2] D. Gregory, G. Shaun, D. Brian and K. Suranga, "Generalization of Machine Learning Approaches to Identify Notifiable Conditions from a Statewide Health Information Exchange," AMIA Jt Summits Transl Sci Proc, 152–161, 2020 .

[3] M. Emmanuel, C. Milena, H. Vincent and D. Thomas, "Deep learning workflow in radiology: a primer," Insights Imaging, 2020.

[4] Ayan, E., & Ünver, H. M. (2019, April). Diagnosis of pneumonia from chest X-ray images using deep learning. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-5). Ieee.

[5] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. Diagnostics, 10(6), 417.

[6] Sirazitdinov, I., Kholiavchenko, M., Mustafaev, T., Yixuan, Y., Kuleev, R., & Ibragimov, B. (2019). Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database. Computers & electrical engineering, 78, 388-399.

[7] Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ... & Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. Applied Sciences, 10(9), 3233.

[8] Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021). Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. Journal of medical systems, 45(7), 1-10.

[9] "https://data.mendeley.com/datasets/rscbjbr9sj/3".

[10] T. Boyadzhiev, S. Tsvetanov and S. Dimitrova, "Comparison of Auto-Encoder Training Algorithms," remote sensing, Proceedings of the IHIET 2021 Conference, 2021.

[11] R. O, F. P and B. T, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical image computing and computer-assisted intervention, 2015.

[12] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[13] Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256). JMLR Workshop and Conference Proceedings.