

Multi-View CNN with MLP for Diagnosing Tuberculosis Patients Using CT Scans and Clinically Relevant Metadata*

Abdela A. Mossa¹, Abdulkemim M. Yibre² and Ulus Çevik³

¹ Department of Computer Engineering, Faculty of Engineering, Çukurova University, 01330 Sarıçam, Adana, Turkey (Email:abdela4u@gmail.com)

²Department of Computer Engineering, Faculty of Engineering, Konya Technical University, 42250, Selçuklu, Konya, Turkey (Email:abdukerimm@selcuk.edu.tr)

³Department of Electrical and Electronics Engineering, Faculty of Engineering, Çukurova University, 01330 Sarıçam, Adana, Turkey (Email:ucevik@cu.edu.tr)

Abstract. We propose a hybrid approach of multi-view convolutional neural networks with Multi-Layer Perceptron to generate an automatic medical CT report and evaluation of the severity stage of Tuberculosis patients, trained and evaluated on 335 chest 3D CT images and available metadata provided by ImageCLEF2019 organizers for the participants of tuberculosis computation track. Transfer learning and data augmentation techniques were applied to avoid over fitting and enhance performance of the model. Our multi-view CNN approach comprises the decomposition of the 3D CT image into 2D axial, coronal and sagittal slices and converting them to PNG format as preliminary to training. At the first stage, coronal and sagittal slices were used to train the CNN classifier using pre-trained AlexNet. In the second stage, MLPs were trained using features extracted during stage one alongside with the provided metadata. Our results ranked 6th and 4th, with an AUC of 0.763 in predicting whether the severity stage is High or Low, and mean AUC of 0.707 in detecting whether left and right lungs are affected or not, detecting the absence or presence of calcifications, caverns, pleurisy and lung capacity decrease, respectively.

Keywords: Tuberculosis Detection, Severity Score, Automatic CT Report, Convolutional Neural Network, Deep Learning, Multi-Layer Perceptron, Medical Imaging Analysis.

1 Introduction

About 130 years after its discovery, Tuberculosis (TB) is one of the 10 leading causes of death in the world. In 2017 alone, TB caused an estimated 1.3 million deaths and around 10.0 million people developed TB disease. With early diagnosis and proper

* Copyright (c) 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2019, 9-12 September 2019, Lugano, Switzerland.

treatment, each year it is possible to prevent millions of people with TB from death [1].

Advanced medical imaging technologies like Computed Tomography (CT), when used by expertise radiologists with the help of Computer Aided Detection (CAD) software, can detect subtle alterations in the lung tissue to correctly identify and diagnose the disease [2]. But despite many advances in both diagnosis and treatment, the application of TB diagnosis remains one of highest causes of mortality from any infectious cause in the world [3] which shows still challenges in detection and treatment of TB are ahead.

It has been reported in [4, 5] that, there is a relative lack of expert radiologists in many TB burden countries, which may impair screening effectiveness and delays the diagnosis results. In India, an average TB patient is diagnosed after a delay of nearly 2 months [6] and overall, a true negative rate as high as 30 % and a false positive rate of up to 15 % has been reported in radiology [2]. Evidently, there is an unmet need to fully automatic CAD for TB diagnoses which is efficient, facilitates earlier detection of disease and save significant health care costs. Hence to tackle these problems, ImageCLEF [7] has presented an evaluation campaign that welcomes researchers around the world to participate in ImageCLEFmed Tuberculosis2019 [8] task for the third consecutive year which comprises two subtasks: Severity Scoring (SVR) and CT Report (CTR). The challenge is based on the 3D CT scans of patients with TB along with the provided automatic report of the patients. The aim of SVR subtask is to assess the severity of each TB patient and classify either to “LOW” (critical) or “HIGH” (very good). The aim of the CTR subtask is to generate an automatic medical report based on the status of Left and right lungs, presence or absence of calcifications, caverns, pleurisy and capacity of lung.

Deep learning approaches [9]– in particular, deep convolutional neural networks (CNNs) have been shown to be successful on a large variety of computer vision and image analysis tasks [10–13] and recently they have also been broadly applied, and are in the infant stage to the medical imaging field. In this paper we provide an artificial intelligence-enhanced CAD technology, which is a fully-automatic hybrid model of CNN and Multi-Layer Perceptron (MLP) for diagnosing people with TB. The inputs to the CNN architecture that we used are sagittal and coronal view point slices. Henceforth, the CNN architecture we used in this paper is named as Multi-View CNN.

This paper is organized as follows. In section 2, a brief overview of the ImageCLEFmed Tuberculosis2019 subtasks, datasets, and preprocessing steps are described in detail. In section 3, we discuss Multi-View CNN and MLP models. The results obtained using our approaches in the two subtasks are shown in Section 4. Finally, Section 5 concludes our participation in this challenge

2 Data-Preprocessing

The training and test datasets provided by the ImageCLEFmed Tuberculosis2019 task organizers consist of 335, for training 218 and 117 for testing, chest CT volumetric

scans of people with TB which are stored in NIFTI file format. The sizes of all volumetric scans are $512 \times 512 \times s$, where image length and width are 512 and s indicates number of slices in the axial plane which varies from 50 to 400. In addition, for all patients the provided datasets also includes automatic extracted masks of the lungs and clinically relevant metadata. However, we do not use these provided segmentation masks in our work.

Two binary classification subtasks of the TB task were proposed by the organizers: i) Severity scoring, ii) CT report. The two sub tasks share the same datasets. For both sub tasks, we split the provided training dataset into training and validation of 174 and 44 volumetric scans, respectively. The validation data was selected from the training using stratified random sampling to avoid bias and ensure that proportional number of positive and negative labels were present in each set. In addition, it allowed us for tuning the hyperparameters and selects the best model for later use to evaluate the test dataset.

We preprocess the NIFTI images for making it compatible for transfer learning using AlexNet [13] and it passes through different stages. First we reconstruct all the 3D CT scan in all of the 3 planes: 2D Axial, sagittal and coronal slices were extracted from each patient’s NIFTI volume data files, then rescaled the image intensity pixel of each slice so that the actual minimum intensity value is mapped to 0 and the actual maximum intensity value is mapped to 255, which is the standard range for PNG images. To avoid processing the background which does not contain any chest tissue and tackle the limited memory of GPUs constrain, some slices at the beginning and end of sagittal and coronal views from each volume scan were discarded, resulting 400 slices (200 sagittal and 200 coronal) from each 3D CT scan. Next, depending on the shape size of each extracted sagittal and coronal slices, cropping or padding was used to rescale each slice to 224×112 pixels.

We curate three separate datasets with all slices having the same shape size, 224×224 , by concatenating sagittal and coronal slices which are on the same position. The first dataset was created by merging sagittal slices on the left and coronal slices on the right and the second dataset was created by merging coronal slices on the left and sagittal slices on the right, reducing the number of slices from 400 (224×112) to 200 (224×224) for a single 3D CT scan. We then convert each merged slices to Portable Network Graphics (PNG) format and normalize to have zero mean and unit variance. The third dataset contains combination of dataset one and two. All the three datasets were used for training our models and only the first dataset for validating and testing the trained models.

All the preprocessing steps were done using the python programming language and NiBabel package [14] . The example of reconstructed, scaled, and merged PNG images displayed using Python and Matplotlib Python package is shown in fig 1. In this work, Axial slices not thoroughly investigated but didn’t improve the performance when we tried hence they were not used in final results.

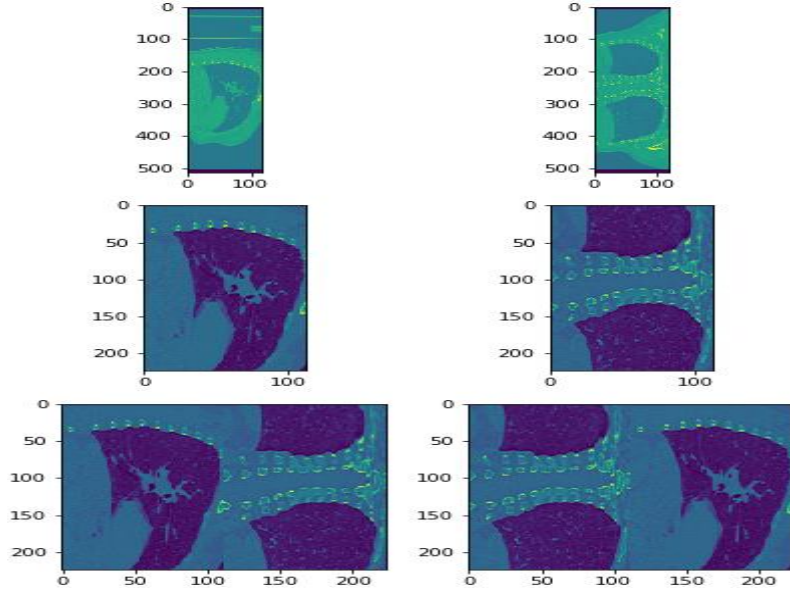


Fig. 1. Example of CT volumetric scan preprocessing stages. First row: sagittal slice; coronal slice; second row: sagittal resized; coronal resized; third row: merged-sagittal on the left and coronal on the right; merged-coronal on the left and sagittal on the right.

3 Model

Problem definition. For each 3D chest CT scans of TB patients, we assign 7 binary labels: high severity (scores 1, 2 and 3)/low severity (scores 4 and 5), left lung affected/not affected, right lung affected/not affected, presence/absence of calcifications, presence/absence of caverns, presence/absence of pleurisy and lung capacity decrease/not decrease. Our goal is to develop 7 similar automatic models to predict the 7 labels for each patient.

Model architecture. Inspired by [15, 16], we developed a hybrid architecture using CNN and MLP. The overall architecture consists of two core modules: (i) transfer learning using pre-trained CNN AlexNet architecture with little modification that maps 2D-slices of a patient to probability prediction as a deep feature extractor for each of the 7 binary classification problems, and (ii) MLP- a standard machine learning classifier that takes deep learning features obtained from the Multi-view CNN model and available metadata as input to display the final TB diagnosis results. See fig 2 for a schematic representation of the experimental set up of our hybrid model.

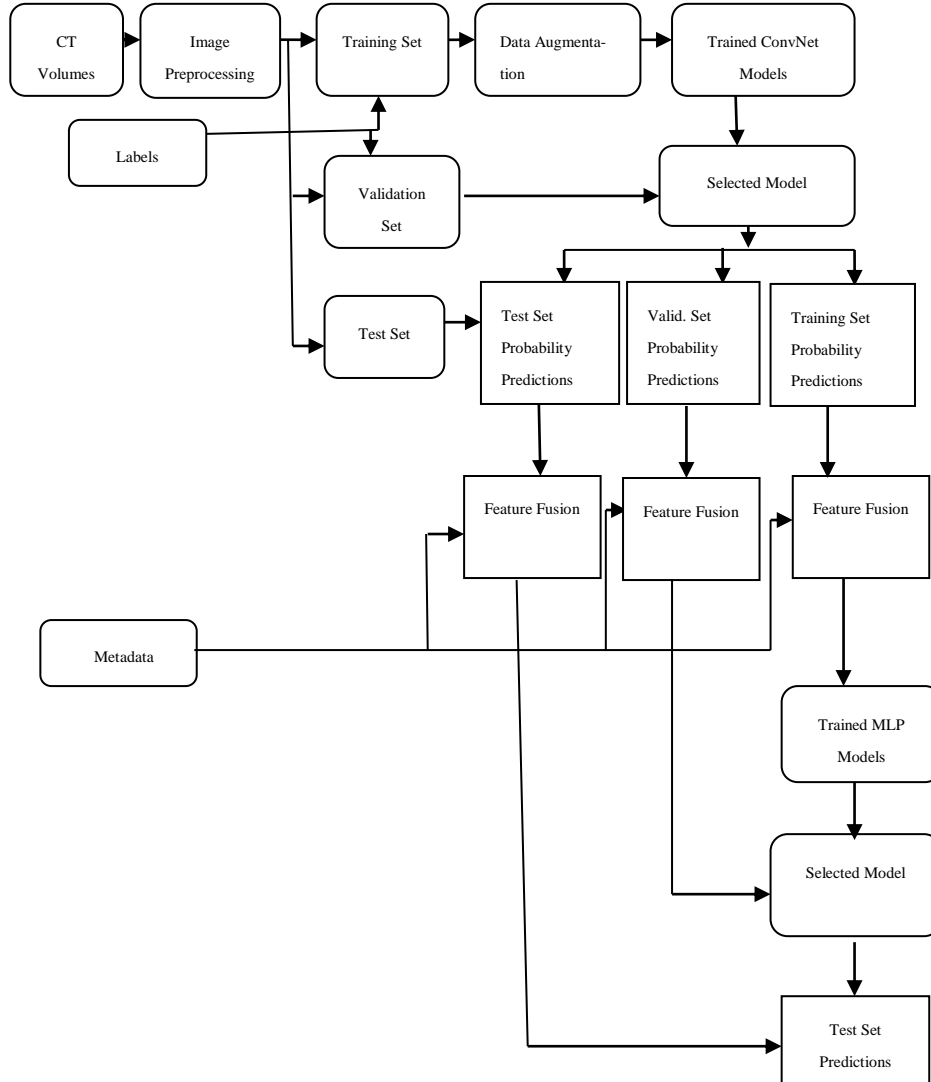


Fig. 2. A schematic representation of the proposed hybrid model for SVR score assessment and CT report generation by using Multi-View CNN to conduct feature learning, and MLP classifier for final prediction using learned features and available metadata. We use this experimental set up for all of the 7 binary classification problems of the two subtasks explained in the previous section.

Multi-View CNN. The trained Multi-view CNN architecture for extracting deep features were 2D convolutional neural networks implemented using Python programming language and PyTorch [17]: open source deep learning platform with medium level abstraction between Tensorflow and Keras. The architecture takes

stacked 2D slices of 3D CT scan as input, with three channels corresponding to RGB, and outputs a probability. As shown in fig 3 below, the overall network architecture consists of three core parts: (i) the convolutional base of pre-trained AlexNet which is a state of the art deep learning model trained on ImageNet database which has 1.2 million high-resolution images belonging to 1000 categories, (ii) global average pooling and max pooling layers on top of the convolutional base applied across the spatial dimensions to reduce features obtained from the convolutional base, and (iii) final dense layer with sigmoid activation function that outputs a probability binary prediction for each subtasks problems.

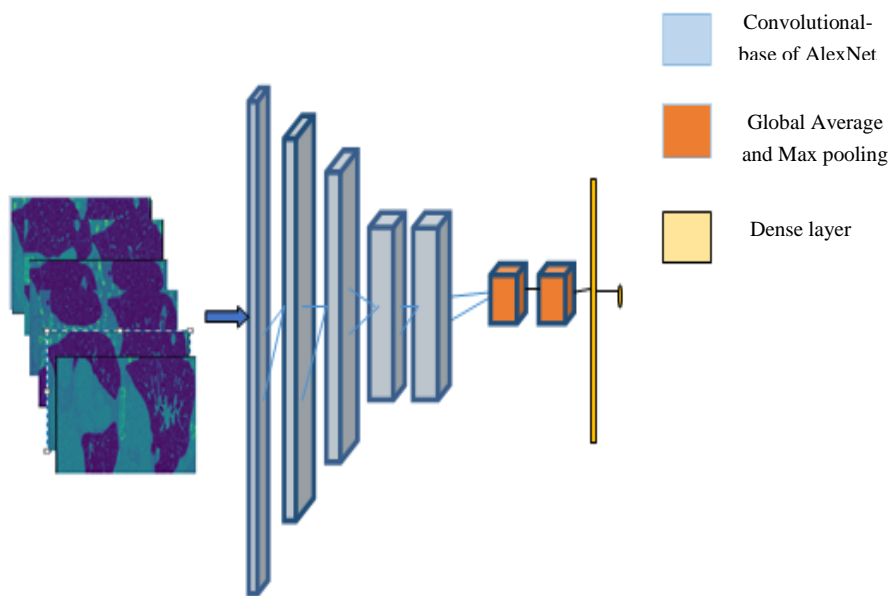


Fig. 3. Multi-View CNN architecture. It takes a stacked $s \times 3 \times 224 \times 224$ dimension pre-processed PNG images of each 3D CT volume of a patient as input and outputs a classification prediction for each binary classification problems where s is the number of merged slices of sagittal and coronal views for a single 3D CT volume of a patient.

We use back propagation algorithm for training and the binary cross-entropy loss function along with Adam optimizer used for optimizing the model using a learning rate of 10^{-5} . Furthermore, during training we used data augmentation techniques to increase the diversity of the data samples to avoid the behavior of overfitting due to the small size of the training dataset. We apply common augmentation techniques such as randomly rotate between 25 and -25 degree, and horizontal flipping to create new images. We did not use any data augmentation techniques during test and validation time.

We trained the network three times for each of the 7 binary classification problems, one for each dataset created in the pre-processing stage, resulting three different deep feature outputs for each patient scanned volume images. During training the MLP

classifier, the combination of the three deep features with the metadata achieved better results than either of them alone with the metadata.

MLP-Classifier. Once we complete training the Multi-View CNN for the three datasets, using MLP classifier we can get the medical diagnosis report of a TB patient, i.e. SVR prediction score and CTR report, using the 3 deep features extracted at the last layer of our CNN architecture and clinically relevant available patients' information metadata. We used Weka data mining tool to perform training and testing the MLP.

4 Results

In addition to the main tests executed using Multi-view CNN and MLP, extra experiments were also tested using Multi-view CNN and various machine learning methods, like Naïve Bayes, Random Forest and Random Tree but all of them were outperformed by MLP. In addition, we observed the performance of Multi-View CNN with and without data augmentation and the model performed better when we used proper number of additional augmented data but performed worse when we tried to use more due to over fitting. We also tried to investigate the performance of the Multi-layer Perceptron architecture by increasing and decreasing the number of hidden layers and nodes in each hidden layers but we got better performance when we used the default architecture of MLP by Weka data mining tool and assigning 0.001 learning rate, 0.2 momentum and 725 epochs.

Table 1. SVR-Severity scoring results of the participant groups.

| Group name | AUC | Accuracy | Rank |
|-----------------------|---------------|---------------|----------|
| UIIP_BioMed | 0.7877 | 0.7179 | 1 |
| UIIP | 0.7754 | 0.7179 | 2 |
| HHU | 0.7695 | 0.6923 | 3 |
| HHU | 0.7660 | 0.6838 | 4 |
| UIIP_BioMed | 0.7636 | 0.7350 | 5 |
| CompElecEngCU | 0.7629 | 0.6581 | 6 |
| San Diego VA HCS/UCSD | 0.7214 | 0.6838 | 7 |
| San Diego VA HCS/UCSD | 0.7214 | 0.6838 | 8 |
| MedGIFT | 0.71 | 0.641 | 9 |
| San Diego VA HCS/UCSD | 0.7123 | 0.6667 | 10 |

Top ten rankings taken from the results provided by the organizer of ImageCLEFmed Tuberculosis2019 for both subtasks are shown in the table 1 and 2. A total of 10 runs could be submitted in each ImageCLEF2019 TB subtasks but due to the limited time constrain we have submitted only once (indicated in bold in Table 1

and 2) and our team ranked 6th and 4th in SVR and CTR subtasks, respectively. The rank of the results is shown in terms of AUC and Accuracy

Table 2. CTR-CT report results of the participant groups.

| Group name | Mean AUC | Min AUC | Rank |
|-----------------------|---------------|---------------|----------|
| UIIP_BioMed | 0.7968 | 0.6860 | 1 |
| UIIP_BioMed | 0.7953 | 0.6766 | 2 |
| UIIP_BioMed | 0.7812 | 0.6766 | 3 |
| CompElecEngCU | 0.7066 | 0.5739 | 4 |
| MedGIFT | 0.6795 | 0.5626 | 5 |
| San Diego VA HCS/UCSD | 0.6631 | 0.5541 | 6 |
| HHU | 0.6591 | 0.5159 | 7 |
| HHU | 0.6560 | 0.5159 | 8 |
| San Diego VA HCS/UCSD | 0.6532 | 0.5904 | 9 |
| UIIP | 0.6464 | 0.4099 | 10 |

5 Conclusion

In this paper, we investigate the use of a combination of pre-trained CNN and MLP classifier to diagnose people with TB and our results show a promising result in ImageCLEF 2019 TB evaluation track. Due to the limited time constraints and availability of limited computational power, we did not use extracted masks of the lungs, axial slices, and some coronal and sagittal slices at the beginning and end even though we expect that it will improve the result. Ultimately we would like to address these issues in the future.

Acknowledgments

This work was supported by the research fund of the Çukurova University, Project Number: 10683

References

1. Global tuberculosis report 2018:World Health Organization; Licence: CC BY-NC-SA 3.0 IGO.
2. Dicente Cid, Y.: Lung Tissue Analysis : From Local Visual Descriptors To Global Modeling, <https://nbn-resolving.org/urn:nbn:ch:unige-1113942>, (2018). <https://doi.org/10.13097/archive-ouverte/unige:111394>.
3. Bomanji, J.B., Gupta, N., Gulat, P., Das, C.J.: Imaging in tuberculosis. Cold Spring Harb. Perspect. Med. 5, 1–24 (2015). <https://doi.org/10.1101/cshperspect.a017814>.
4. Hoog, A.H. van't, Meme, H., van Deutekom, H., Mithika, A., Olunga, C., Onyino,

- F., Borgdorff, M.: High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. *Int. J. Tuberc. Lung Dis.* 15, 1308–1314 (2011). <https://doi.org/https://doi.org/10.5588/ijtld.11.0004>.
5. Melendez, J., Sánchez, C.I., Philipsen, R.H.H.M., Maduskar, P., Dawson, R., Theron, G., Dheda, K., Van Ginneken, B.: An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci. Rep.* 6, 1–8 (2016). <https://doi.org/10.1038/srep25265>.
 6. Sreeramareddy, C.T., Qin, Z.Z., Satyanarayana, S., Subbaraman, R., Pai, M.: Delays in diagnosis and treatment of pulmonary tuberculosis in India: A systematic review. *Int. J. Tuberc. Lung Dis.* 18, 255–266 (2014). <https://doi.org/10.5588/ijtld.13.0585>.
 7. Dicente Cid, Y., Liauchuk, V., Klimuk, D., Tarasau, A., Kovalev, V., Müller, H.: Overview of ImageCLEFtuberculosis 2019 - Automatic CT-based Report Generation and Tuberculosis Severity Assessment. In: CLEF 2019 Working Notes. CEUR Workshop Proceedings (CEUR-WS.org), ISSN 1613-0073, <http://ceur-ws.org/Vol-2380/>, Lugano, Switzerland (2019).
 8. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A. Ben, Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.-T., Piras, L., Riegler, M., Tran, M.-T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., Herrera, A.G.S. de, Garcia, N., Kavallieratou, E., Blanco, C.R. del, Rodríguez, C.C., Vasilopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. In: Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), Lugano, Switzerland, LNCS Lecture Notes in Computer Science, Springer (September 9-12 2019).
 9. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature.* 521, 436 (2015).
 10. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. 1–14 (2014).
 11. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88 (2017). <https://doi.org/https://doi.org/10.1016/j.media.2017.07.005>.
 12. Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mouggiakakou, S.: Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans. Med. Imaging.* 35, 1207–1216 (2016). <https://doi.org/10.1109/TMI.2016.2535865>.
 13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira, F., Burges, C.J.C., Bottou, L., and Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems* 25. pp. 1097–1105. Curran Associates, Inc. (2012).
 14. Matthew Brett, Michael Hanke, MARC-ALEXANDRE CÔTÉ, Paul McCarthy, Chris Cheng, B.C.: NiBabel- Access a cacophony of neuro-imaging file formats, <https://nipy.org/nibabel>. <https://doi.org/10.5281/zenodo.2530243>.
 15. Geras, K.J., Wolfson, S., Shen, Y., Wu, N., Kim, S.G., Kim, E., Heacock, L., Parikh, U., Moy, L., Cho, K.: High-Resolution Breast Cancer Screening with Multi-View

Deep Convolutional Neural Networks. 1–9 (2017).

16. Bien, N., Rajpurkar, P., Ball, R.L., Irvin, J., Park, A., Jones, E., Bereket, M., Patel, B.N., Yeom, K.W., Shpanskaya, K., Halabi, S., Zucker, E., Fanton, G., Amanatullah, D.F., Beaulieu, C.F., Riley, G.M., Stewart, R.J., Blankenberg, F.G., Larson, D.B., Jones, R.H., Langlotz, C.P., Ng, A.Y., Lungren, M.P.: Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med.* 15, 1–19 (2018). <https://doi.org/10.1371/journal.pmed.1002699>.
17. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in PyTorch. In: 31st Conference on Neural Information Processing System. , CA, USA (2017).