Automated Chest X-ray Disease Diagnosis Using Deep Learning Approaches

Xinxin Zhai, Fei He

Georgia Institute of Technology, Atlanta, GA, USA

Motivation

The chest X-ray is the most widely performed radiography imaging test for patient health monitoring and disease diagnosing such as pneumonia, heart failure, lung cancer, tuberculosis, sarcoidosis, and lung tissue scarring, etc.[1]. However, their proper usage in healthcare services depends largely on the correct interpretation of medical information hidden in these images, which require lots of professional training for clinicians. Here we propose to use deep learning-based approaches (e.g., convolutional neural network (CNN) architectures) to solve the major obstacle here—accurate classification of chest X-rays with meaningful disease information. We will implement and optimize CNN-based algorithms with hyperparameter tuning and compare the modeling performance with previous studies [2-5]. This practice will greatly benefit disease diagnosis with improved efficiency through automated chest X-ray analysis.

Literature review

A convolutional neural network is a class of deep neural networks that is commonly used for visual imagery analysis [6]. It consists of an input and an output layer as well as multiple in-between hidden layers, which are made of neurons as the basic computation units. Several previous studies [2-5] have applied CNN-based architectures for disease classification in chest X-ray images. In 2017, Wang et al. [2] firstly presented a chest X-ray database named “ChestX-ray8” with 108,948 frontal-view X-ray images for 8 diseases from the associated radiological reports of 32,717 patients. They also employed a unified weakly-supervised Deep Convolutional Neural Network (DCNN) architecture to realize multi-label image classification and disease localization based on this database. Rajpurkar et al. [3] followed their work by introducing a 121-layer CNN model (ChexNet) to improve disease detection accuracy with the new ChestX-ray14 dataset. Similarly, Liu et al. [4] also improved modeling performance of chest X-ray image classification and disease detection by using the same ChestX-ray14 dataset and a new segmentation-based deep fusion network (SDFN) that leverages the higher-resolution information of local lung regions. In 2019, Irvin et al. [5] presented a new CheXpert dataset containing 224,316 chest radiographs of 65,240 patients and designed a new CNN-based disease labeler that captures uncertainties in radiograph interpretation. They evaluated their model performance against board-certified radiologists and concluded that their best model outperformed human professionals in the detection of 5 selected pathologies. These successful applications of CNN-based architectures for chest X-ray image classification and disease labeling highlight the great potential of deep learning approaches in radiography interpretation and encourage our practice here to try to reproduce their successes of deep learning approaches in medical care diagnosis.

Data

Chest X-ray data from NIH Clinic Center are downloaded for the model training and testing [7]. In total, 112,210 chest X-rays of 30,805 patients from multiple visits are included in the dataset. Some patients are labeled with multiple findings, such as “Effusion|Emphysema|Infiltration|Pneumothorax”. In our training, we will use the chest X-rays with multiple labels for each label identified. For example, if a patient’s findings include Emphysema and Infiltration, we will use it for both targets. Overall, there are 141,537 labels generated. 14 pathologies are labeled in the dataset, i.e., Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural\_thickening, Cardiomegaly, Nodule, Mass and Hernia. The distribution of each label is shown below (Table 1). Other information that can be used for segmentation including patient age, gender, and view position. We will test on segmentation schema and find the best performance solution.

**Table 1.** Distribution of identified pathologies.

|  |  |  |  |
| --- | --- | --- | --- |
| Finding Labels | count | Male count | Female count |
| Cardiomegaly | 2,776 | 1,307 | 1,469 |
| Hernia | 227 | 96 | 131 |
| Mass | 5,782 | 3,529 | 2,253 |
| Consolidation | 4,667 | 2,666 | 2,001 |
| Infiltration | 19,894 | 11,427 | 8,467 |
| Effusion | 13,317 | 7,435 | 5,882 |
| Nodule | 6,331 | 3,685 | 2,646 |
| Emphysema | 2,516 | 1,610 | 906 |
| Atelectasis | 11,559 | 6,906 | 4,653 |
| Pleural\_Thickening | 3,385 | 2,042 | 1,343 |
| Pneumothorax | 5,302 | 2,717 | 2,585 |
| Fibrosis | 1,686 | 915 | 771 |
| Edema | 2,303 | 1,204 | 1,099 |
| Pneumonia | 1,431 | 838 | 593 |
| No Finding | 60,361 | 33,922 | 26,439 |
| total identified | 81,176 | 46,377 | 34,799 |
| total | 141,537 | 80,299 | 61,238 |

Approach and Experimental Setup

We will employ Apache Spark, a unified analytics engine for large-scale data management and processing, to process the chest X-ray images listed above on a local standalone Linux-based cluster. Though Spark has internal built machine learning library—MLlib—for practical machine learning pipelines such as classification, regression, clustering, and collaborative filtering, it does not support more sophisticated algorithms including deep neural network architectures like Convolutional Neural Network (CNN) for image processing. Therefore, we are going to integrate TensorFlow—an open source software library developed by the Google Brain team for high performance numerical computation using data flow graphs—with Spark and implement this hybrid platform for our deep learning practice of chest X-ray image classification. The Spark-TensorFlow pipelines will also show great advantages for hyperparameter tuning when building CNN models in a distributed manner, which allows manipulation at scale and improves modeling speed and accuracy.

Specifically, we are going to use TensorFlow with Spark Deep Learning pipelines (pyspark from Databricks) on Python 2.7 to transform X-ray images on numeric features. Firstly, we will use utility functions in Spark Deep Learning pipelines to load a part of the X-ray images for model training purpose into a Spark DataFrame and decode them automatically in a distributed way. Then we produce a model using the Transfer Learning technique for image classification. We will combine InceptionV3 (a convolutional neural network trained for image classification) in Spark and logistic regression. The DeepImageFeaturizer will automatically peels off the last layer of a pre-trained neural network and uses the output from all the previous layers as features for the logistical regression part. We will also use the rest of the X-ray images to test performance of the trained deep learning CNN models. We will follow the previous studies [2,3] to use Area-Under-Curve (AUC) and F1 scores as main metrics for the model performance evaluation.

Timeline

Below is our proposed timeline for completing this project.

**Table 2.** Timeline**.**

|  |  |  |
| --- | --- | --- |
| Time | Task | Deliverable |
| March 4 | Gather data and do exploratory data analysis | Descriptive statistics |
| March 10 | Define cohort, target, and features | Get a clear definition of target and features |
| March 16 | Data cleaning and processing | Remove bad records |
| March 20 | Test the data using multiple modeling techniques | Get quick models and make sure the target and features are reasonable |
| March 24 | Refine models | Get refined models |
| March 28 | Evaluate the model candidates on the performance metrics | Compare model performance |
| April 1 | Interpret the results from the final model (e.g., show predictive  features, compare to literature in terms of finding, visualization) | Write interpretation and generate figures for better presentation |
| April 5 | Write first draft for team discussion | First draft |
| April 6 | Team edits the draft | Submit project draft |
| April 15 | Add additional analysis such as visualization and literature comparison to refine the paper | Edited paper |
| April 20 | Make presentation | Deck for presentation |
| April 24 | Final paper, codes, and deck submission | Final paper, codes, and deck |

References

1. <https://www.nhlbi.nih.gov/health-topics/chest-x-ray>
2. Wang X., Peng Y., Lu L., Lu Z., Bagheri M., et al., ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. IEEE CVPR 2017, pp. 2097-2106, [10.1109/CVPR.2017.369](https://arxiv.org/ct?url=https%3A%2F%2Fdx.doi.org%2F10.1109%252FCVPR.2017.369&v=8fa18705), arXiv:1705.02315 [cs.CV], 2017.
3. Rajpurkar P., Irvin J., Zhu K., Yang B., Mehta H., et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225, 2017.
4. Irvin J., Rajpurkar P., Ko M., Yu Y., Ciurea-Ilcus S., et al., CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. AAAI 2019, arXiv:1901.07031v1 [cs.CV], 2019.
5. Liu H., Wang L., Nan Y., Jin F., and Pu J. SDFN: Segmentation-based Deep Fusion Network for Thoracic Disease Classification in Chest X-ray Images. arXiv:1810.12959 [cs.CV], 2018.
6. https://en.wikipedia.org/wiki/Convolutional\_neural\_network
7. <https://nihcc.app.box.com/v/ChestXray-NIHCC/folder/36938765345>
8. Gardner RM, Golubjatnikov OK, Laub RM, Jacobson JT, Evans RS. Computer-critiqued blood ordering using the HELP system. Comput Biomed Res 1990;23:514-28.