The University of Melbourne - School of Computing and Information Systems

Utilising deep learning-based chest X-ray diagnoses with public data

Team 13 Tuan Khoi Nguyen (1025294) Xiaochen Hou (1067898) Ivy Liang (1397138) Clemence Mottez (1486585) Qichi Liang (1392005)

1 Background & Problem Definition (104 words)

In current global health technology, respiratory diagnostics have witnessed the use of machine learning (ML) for chest X-ray (CXR) analysis [1, 2], which reduces diagnosis time significantly for practitioners [3]. However, current potential still remains large, as they only rely on imagery data without considering external factors such as policies or intervention procedures [4].

With great availability of public data and computer vision models, this research endeavours to utilise ML further for CXR diagnostics, by combining imagery and non-imagery data. This will not only aim to retain deep learning's fast operation speed, but also improve prediction correctness with transfer learning concept, maximising efficiency of ML in aiding practitioners.

2 Data Source & Cohort Processing (209 words)

2.1 Data source

Subjects The database used for prediction in this project is Medical Information Mart for Intensive Care (MIMIC) [5, 6], which is distinguished by its diverse and comprehensive set of de-identified patient data. For this project, we retrieve patients' CXR images along with their admission information in hospital or intensive care unit.

Pre-extraction data In order to reflect the nature of unseen data and avoid overfitting in neural network stage, we avoid using pre-trained weights from MIMIC. Therefore, CheXpert [7], a Stanford's CXR database with 224,316 images will be used for transfer learning training.

2.2 Cohort Extraction

Ground truth Common lung diseases or dysfunctions will be attempted to diagnose in this project: for example pneumonia, tuberculosis, bronchitis, lung cancer. We will use ICD code [8] assigned to each patient as ground truth in the dataset.

Cohort condition As we are evaluating the effect of combining CXR and patient data on diseases, each cohort subject has to satisfy 2 conditions. They need to have at least 1 CXR image, and have at least 1 diagnosed ICD code that belongs to a lung disease or dysfunction.

Cohort features To give the most sufficient amount of information, respiratory-related measurements such as blood pressure or pulse oximetry in the form of time series are taken, and will be processed into time series characteristics, both categorical (identifying if an event happened), or numerical (critical values such as mean or maximum).

3 Pipeline & Expectations (316 words)

3.1 Processing pipeline

After retrieving the data needed for the project, the next part of the project will perform a sequence of ML steps on the data, and subsequently evaluate their performance at an in-depth level, critically assessing strengths and weaknesses of the proposed approach. A descriptive diagram is shown in Figure 1.

Transfer Learning & Fusion The proposed system is based on the concept of transfer learning. From a deep learning model pre-trained on common chest x-ray datasets, latent representations of the image can be extracted [9]. This provide learning information without costing training time and allows learning process to then infuse external features outside imagery data.

Models: ResNet-CNN & ML Classifier The deep learning procedure will use ResNet, one of the most common deep architectures in computer vision. It incorporates high-level information between different layers, allowing features to be analysed deeply with sufficient amount of information [10], giving the system the ability to provide the best informed prediction. The final stage of the project will then put processed features into a chosen ML classifier model to predict the chance of each disease occurring.

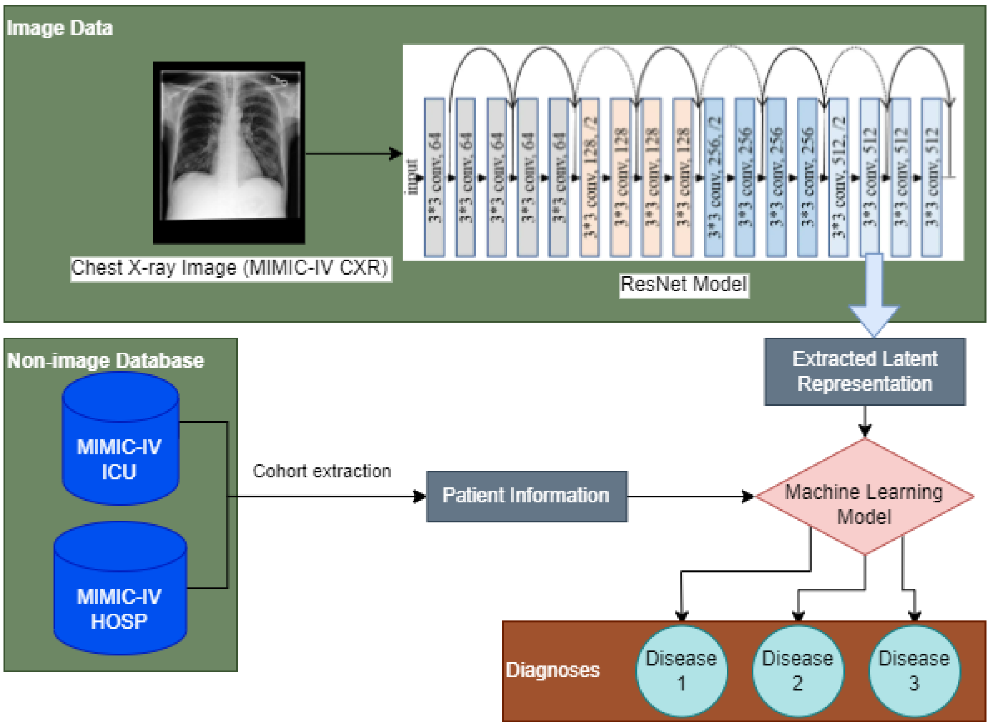

Figure 1: Diagram of the project's suggested model pipeline

3.2 Outcome Evaluation

As we aim to improve diagnosis with additional non-image features, direct comparisons will be made between models with and without them. The former is expected to show improvement in classifying performance.

3.3 Metrics

Performance With the problem being multi-label classification, standard metrics of Precision, Recall, and F1-Score are used to give a comprehensive view on system's ability to avoid Type I and II errors on each disease. AUC-ROC [11] will also be employed, as it is able to evaluate the model's distinguishing power through plotting with different likelihood thresholds.

Feature importance Other than performance metrics, interpretability is also crucial. Therefore, we will perform exploratory data analysis to identify the main features of interest in the system's decision-making, which includes finding correlations and trends between features and predicted labels, and look for key interest regions in the image that the neural network model have extracted for learning using GRAD-CAM, a gradient-based visualisation method [12].

REFERENCES

[1] Yeli Feng, Hui Seong Teh, and Yiyu Cai. Deep learning for chest radiology: A review. *Current Radiology Reports*, 7(8), July 2019.

[2] Matija Soric, Danijela Pongrac, and Inaki Inza. Using convolutional neural network for chest x-ray image classification. In *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE, September 2020.

[3] Hassan K. Ahmad, Michael R. Milne, Quinlan D. Buchlak, Nalan Ektas, Georgina Sanderson, Hadi Chamtie, Sajith Karunasena, Jason Chiang, Xavier Holt, Cyril H. M. Tang, Jarrel C. Y. Seah, Georgina Bottrell, Nazanin Esmaili, Peter Brotchie, and Catherine Jones. Machine learning augmented interpretation of chest x-rays: A systematic review. *Diagnostics*, 13(4):743, February 2023.

[4] Filippo Pesapane, Marina Codari, and Francesco Sardanelli. Artificial intelligence in medical imaging: threat or opportunity? radiologists again at the forefront of innovation in medicine. *European Radiology Experimental*, 2(1), October 2018.

[5] Alistair E. W. Johnson, Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Chih ying Deng, Roger G. Mark, and Steven Horng. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*, 6(1), December 2019.

[6] Alistair Johnson, Lucas Bulgarelli, Tom Pollard, Steven Horng, Leo Anthony Celi, and Roger Mark. Mimic-iv, 2023.

[7] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren, and Andrew Y. Ng. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):590–597, July 2019.

[8] World Health Organization. Icd-10 : international statistical classification of diseases and related health problems : tenth revision, 2004.

[9] Mohamed Berrimi, Skander Hamdi, Raoudha Yahia Cherif, Abdelouahab Moussaoui, Mourad Oussalah, and Mafaza Chabane. COVID-19 detection from xray and CT scans using transfer learning. In *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*. IEEE, March 2021.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

[11] Michael H Ferris, Michael McLaughlin, Samuel Grieggs, Soundararajan Ezekiel, Erik Blasch, Mark Alford, Maria Cornacchia, and Adnan Bubalo. Using ROC curves and AUC to evaluate performance of no-reference image fusion metrics. In *2015 National Aerospace and Electronics Conference (NAECON)*. IEEE, June 2015.

[12] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, October 2017.