Project Report (April 15, 2021)

Ankita Ghosh[1] and Sahil Khose[2]

[1]Research Assistant, ghoshankita0907@gmail.com , CSE, MIT Manipal
[2]Research Assistant, sahilkhose18@gmail.com, ICT, MIT Manipal

ABSTRACT

We discuss the various approaches towards detecting malignant lung cells using Deep Learning based CNNs on histopathological data. Based on the existing literature review, we have consolidated the line of action for approaching the given problem statement.

Related Work

In this section, we first revisit previous works on classification of benign and malignant cells from lung cytological images using deep CNN.

Data pre-processing

- The dataset contains **slides** of cancerous and non-cancerous tissues. **(x x y) dimension tiles** are cropped from the slides.
- Only **tissue parts are extracted** and the background is cleaned using classical image processing techniques like thresholding before **labeling** the cropped images.
- **Data augmentation** is performed on the dataset using various methods which include rotation, shifting, morphological transforms, colour offset and the use of filters.

The dataset is then used for training the Deep Learning Model.

Deep Learning methods

- Nicolas et al: Trained **Inception v3**.
  - They first trained the network to recognize tumor vs normal observed 0.993 AUC.
  - They tested the performance of their approach for distinguishing LUAD and LUSC which is a much harder task observed 0.847 AUC.
  - It was found that fully training the network instead of just fine-tuning it yielded better results.
  - They also trained the network on direct three-way classification: (normal, LUAD, LUSC) and observed the best results at AUC 0.968
- Atsushi Teramoto et al: Trained **VGG-16** as their primary model.
  - They reported comparisons with **GoogleNet**, **ResNet-50**, **DenseNet**.
  - They adopted three types of classification approaches: **patch-based classification**, **case-based classification** and **distribution of malignant cells in microscopic images**.
- Osamu et al: Trained **Inception v3**

Evaluation Metrics

- **Sensitivity**: (false negative rate) Sensitivity = (True Positive)/(True Positive + False Negative)
- **Specificity**: (false positive rate) Specificity = (True Negative)/(True Negative + False Positive)
- **Accuracy**: Percentage of correct classification prediction.
- **AUC (Area Under The Curve)**: AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0s as 0s and 1s as 1s. By analogy, the Higher the AUC, the better the model is at distinguishing between malignant and benign tumors.

Discussion

Considering the advancements in the Deep Learning architectures in the recent years, the dataset can be trained on superior models (better accuracies, fewer parameters) to obtain improved results. Further, the problem statement can be expanded to tasks such as segmentation, detection and localization.

References

1. Janowczyk A, M. A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inf.* DOI: https://doi.org/10.4103/2153-3539.186902 (2016).

2. Teramoto, A. *et al.* Automated classification of benign and malignant cells from lung cytological images using deep convolutional neural network. *Informatics Medicine Unlocked* **16**, 100205, DOI: https://doi.org/10.1016/j.imu.2019.100205 (2019).

3. Iizuka, K. F. K. K. e. a., O. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Sci Rep 10* DOI: https://doi.org/10.1038/s41598-020-58467-9 (2020).

4. Coudray, N. *et al.* Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *bioRxiv* DOI: 10.1101/197574 (2017). https://www.biorxiv.org/content/early/2017/10/03/197574.full.pdf.

5. Pantanowitz, L. *et al.* An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *The Lancet Digit. Heal.* **2**, e407–e416, DOI: https://doi.org/10.1016/S2589-7500(20)30159-X (2020).

12345