An algorithm for classification of ovarian cancer histopathology images and prediction of genetic variants

Seo Jeong Shin¹, MS, Jin Roh², MD, Ph.D, Seng Chan You³, MD, MS, Ho kyun Jeon⁴, Suk-Joon Chang⁵, MD, Hee-Sug Ryu⁶, MD, Jaeg-Hee Kim⁷, MD, Rae Woong Park⁸,⁹, MD, Ph.D

¹Department of Biomedical Science, Ajou University Graduate School of Medicine, Republic of Korea; ²Department of Pathology, Ajou University Hospital, Republic of Korea; ³Department of Biomedical Informatics, Ajou University School of Medicine, Republic of Korea; ⁴Department of Obstetrics and Gynecology, Ajou University School of Medicine, Republic of Korea

"Can we use a microscopic pathology images to recognize cancer-region and predict potential genetic mutation?"

Background & Objective

Ovarian cancer
- Has the highest mortality among gynecologic cancer:
  - 70% of the patients were found with over stage III
- Final diagnosis of tumor malignancy:
  - through histopathologic examination of mass
- Targeted therapy (Olaparib, rucaparib, niraparib):
  - Patient with BRCA1/2 pathogenic mutation can use PARPi as targeted therapy preventing an apoptosis by DNA damage

Let's develop a deep learning algorithm based on genomic CDM for automatic (1) classification of cancer-region and (2) prediction of genetic variants using microscopic pathology images and DNA mutation data of patients with ovarian cancer.

Method & Results

1. Histopathology image and DNA sequence data
   - Whole-slide images and genomic mutation data of patients with high-grade serous carcinoma were obtained.
     1) Ajou University Medical Center (AUMC) → Image & genomic data
     2) The Imaging Archive (TCA) → Image data
     3) The Cancer Genome Atlas (TCGA) → Genomic data
   - The mutation data were converted to genomic-CDM,
     - which has been developing in OHDSI as an extension of OMOP-CDM

2. Mutation status comparison
   - Pathogenic variants important in ovarian cancer.
     - Variants in genes such as TP53, BRCA1, and BRCA2 were compared between datasets using the open source data visualization & analysis tool “GeneProfiler” based on Genomic-CDM.
   - Results were consistent with previous studies.
     - The frequencies of pathogenic mutations in AUMC were similar in TCGA, except for the BRCA1 (8.2-30.3%) (Figure 1).

3. CNN algorithm for classification and prediction
   - The CNN algorithm (Figure 2) developed using the AUMC dataset was externally validated with new samples, TICA dataset.
   - The accuracy in classifying a cancer-region was 0.90 (AUC 0.99), and in predicting BRCA1/2 mutation was 0.66 (AUC 0.71; Figure 3).

4. HeatMap visualization
   - Can our algorithm accurately find the lesions in the entire image even in the absence of the region-assignment by pathologists?
     - The algorithms were learned using the ROI (region of interest) area, which is a part of the overall tissue image.
     - The whole-slide image was cut to the ROI size and then put into the model to classify cancer-region.
     - The corresponding predicted values were used to represent colors at the exact position of the ROI.
     - The results of the HeatMap visualization in the whole image indicated that cancer-region tiles showed high cancer prediction scores compared with the overall region as we expected (Figure 4).

Results: Internal validation (AUMC)
- Classification: Accuracy = 0.901
- AUC = 0.956

Results: External validation (TICA)
- Classification: Accuracy = 0.901
- AUC = 0.956

Figure 2. Convolutional neural network (CNN) model architecture
Figure 3. Accuracy of the cancer classification model and the mutation prediction model.

This work was supported by the Bio Industrial Strategic Technology Development Program (20001234) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea) and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea [grant number: HI16C0922]