Machine Learning Improves the Accuracy of Adequacy Evaluation of Cytology Specimens for Molecular Profiling

Mehrvash Haghighi[1], Max Signaevsky[2], Marcel Prastawa[1], Adnan Hasanovic[2], Brandon Veremis[3], Roshanak Alialy[4], XiangFu Zhang[2], Baidarbhi Chakraborty[2], Gerardo Fernandez[2]

[1]Mount Sinai Hospital
[2]Icahn School of Medicine at Mount Sinai
[3]Mount Sinai Hospital Icahn School of Medicine
[4]Mount Sinai St. Luke's Roosevelt Hospital

Background:
Molecular testing has become standard of care in the diagnosis and therapeutic management of lung carcinomas. Adequate sampling of tumor tissue is crucial for accurate molecular profiling. The percentage of tumor cells in relationship with the benign cellular component of the specimen is the main factor in the determination of sample suitability for accurate genetic testing. Performing the gene sequencing on inadequate samples delay the result and impose a cost on the molecular laboratory. The overestimation of cellularity could lead to false-negative results and withholding therapies from patients. The underestimation of cellularity with negative results could lead to unnecessary repeat biopsies. The inter-rater reliability study has shown very little concordance between raters in assessing the percentage of malignant cells in the specimen, being a motivational factor for the development of an AI approach.

Design:
The percentage of tumor cells on H&E slides of 17 aspirate of lung lesions and 27 pleural fluids, all containing lung adenocarcinomas, were scored by a group of four pathologists. We trained a fully convolutional neural network that detects malignant and benign components on the specimen. We used the SegNet architecture, with modifications that add squeeze-excitation layers. The network is trained using 12000 images of size 512 x 512 pixels extracted from 44 slides. We randomly assign 4 slides as a test data set for measuring the performance of our neural network, obtaining a Dice overlap score 0f 0.73 on annotated 2800 images of size 512 x 512 pixels. The neural network enables the generation of automatic scoring, for example using the ratio of malignant nuclei over total epithelial cells nuclei.

Results:
We use Krippendorff's alpha to measure inter-rater agreement and obtained alpha of 0.275 signifying high variability between raters that motivate our development of an automatic algorithm. The correlations (Spearman's, *Rho*) between the score based on the network detection and estimations by four different pathologists were *Rho*1 =.228, *p*=.158; *Rho*2 =.683, *p*<.001; *Rho*3 =682, *p*<.001; and *Rho*4 =.350, *p*=.027, showing that estimations by two pathologists out of four were in significant correlation with the AI-derived score.

Conclusion:
The artificial intelligence approach could reliably improve the accuracy of evaluation of malignant component on the cytological specimen without effects of inter- and intra-observer variability on molecular testing, which in turn improve the validity of the molecular testing result and patients' outcome.