Machine learning for glioblastoma screening from histopathology whole slide imaging

Cheung EYW\textsuperscript{1}, Li ASM\textsuperscript{1}, Wu RWK\textsuperscript{2}, Chu ESM\textsuperscript{1}

\textsuperscript{1}corresponding author: evacheung@twc.edu.hk, \textsuperscript{2}School of Medical and Health Sciences, Tung Wah College, Hong Kong, \textsuperscript{3}Department of Biological and Biomedical Sciences, School of Health and Life Sciences, Glasgow Caledonian University, Glasgow, Scotland, United Kingdom

Background
Glioblastoma (GBM) is one of the most common malignant primary brain tumours, which accounts for 60-70\% of all glioma. Conventional diagnosis and post-operation treatment plan for glioblastoma is mainly based on the identification of tumor cells on haematoxylin and eosin stained (H&E) histopathological slides by both experienced medical technologist and pathologist. The recent development of digital whole slide scanners allows the histopathological image analysis feasible. This study aimed to build an image feature based model using histopathology whole slide images to differentiate patients with glioblastoma (GBM) from healthy control (HC).

Method
628 H&E images were collected from 30 GBM patients, obtained from public database the Cancer Image Archive. Another 554 images were collected from 30 health control (HC). 22 Gray level Co-occurrence Matrix (GLCM) image features were retrieved from the collected images by an in-house developed software. Both the GBM and HC groups were divided into a training and a validation group on 70:30 basis. A random forest algorithm was used to build the model using training set, and validated by the validation set by R software.

Results
The receiver operating characteristics (ROC) curve showed that the average class and overall accuracy achieved 94.6\%. The area under the curve (AUC) was 0.98, with sensitivity of 94.1\% and specificity of 95.1\%.

Model Performance

\begin{tabular}{|l|c|}
\hline
Overall Accuracy & 94.6\% \\
Sensitivity & 94.1\% \\
Specificity & 95.1\% \\
Area Under ROC & 98.0\% \\
\hline
\end{tabular}

Conclusion
The VMAT in this study showed more favorable DVP in the heart and the LAD for early stage left breast cancer when compared to 3D-CRT. The features of dose distribution can be applied in precision medicine in oncology.