Revolutionizing Computer Vision Algorithms in Cancer Pathology: The Use of Comprehensive Toolkits to Overcome Machine Learning Obstacles in the Digital Pathology Field

Chan, Benjamin (School: Bronx High School of Science)

Applications of machine learning (ML) to the cancer pathology field will lead to earlier and more accurate cancer diagnosis. Cancer must be treated in its early stages for the treatment to be effective. Laboratory staining practices, which result in poor color quality in pathology images, have led to the lack of standardized data for training of AI algorithms. Every currently existing method of color normalization, an algorithm designed to computationally correct poor image quality, has only limited applications and none of the proposed methods are suitable for ML applications. Experts have not found solutions to these problems and were ineffective in solving today's image standardization problems. This study proposes a color normalization "toolkit" consisting of tools to solve modern-day problems in image standardization and color normalization workflows, as well as an objective Validation Framework to mathematically assess the effectiveness of the proposed toolkit. Using the Validation Framework, I determined that, collectively, these methods lead to objectively better results than color normalization without the use of tools. The proposed methods resulted in improvements ranging from 3-5% (for a minor adjustment limited by computational constraints) to as much as 89%. If combined, these tools achieve even larger improvements in image standardization, between 80-90%. By paving the way for ML-based cancer diagnosis, this constitutes a major breakthrough for experts in the field and addresses existing limitations in image standardization. My research will provide a strong pathway to produce high quality images for ML applications, which can lead to earlier, more accurate cancer diagnoses.