Siemens Competition
Math : Science : Technology
National Finalist

Names: David Zhu & Evani Radiya-Dixit
High School: The Harker School
Mentor: Andrew Beck, Beth Israel Deaconess Medical Center
Project Title: *Automated Classification of Benign and Malignant Proliferative Breast Cancer Lesions*

Misclassification of breast lesions can result in either cancer progression or unnecessary chemotherapy. Automated classification tools are seen as promising second opinion providers in reducing such errors. We have developed predictive algorithms that automate the categorization of breast lesions as either benign usual ductal hyperplasia (UDH) or malignant ductal carcinoma in situ (DCIS). From diagnosed breast biopsy images from two hospitals, we obtained 392 biomarkers using Dong et al.'s (2014) computational tools for nuclei identification and feature extraction. We implemented six machine learning models and enhanced them by reducing prediction variance, extracting active features, and combining multiple algorithms. We used the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for performance evaluation. Our top-performing model, a Combined model with Active Feature Extraction (CAFE) consisting of two logistic regression algorithms, obtained an AUC of 0.918 when trained on data from one hospital and tested on samples of the other, a statistically significant improvement over Dong et al.'s AUC of 0.858. Pathologists can substantially improve their diagnoses by using it as an unbiased validator. In the future, our work can also serve as a valuable methodology for differentiating between low-grade and high-grade DCIS.