DOGMA (Dogs' Oncological Genomic Metabolite Analyzer): Diagnosis of CMTs in Dogs Using a Machine Learning Model

Wang, Asher (School: Holicong Middle School)

Canine mammary tumors (CMTs) represent the most common neoplasia in female dogs. Histopathology analysis is considered the gold standard for classification but is invasive and does not always lead to diagnosis. This study aimed to create a novel AI-powered diagnostic model leveraging urinary metabolite profiling to diagnose malignant masses and copilot tumor grading of CMTs. This study employed a large-scale dataset from MetaboLights Repository containing 55 tumor-specific urinary metabolites in 156 dogs collected using proton nuclear magnetic resonance spectroscopy. Various commonly used machine learning algorithms, such as Random Forest (RF), were implemented to build and optimize the models. Results were reported in confusion matrix form in accordance with STARD-AI guidelines. Additionally, feature engineering was utilized to determine each variable's discriminative power. Overall survival (OS) by metabolic grade was compared using log-rank-test. The Malignant Versus Benign (MVB) RF model performed the best (94.74% accuracy, sensitivity, and specificity). Tumors were graded with 82.35% overall accuracy, 72.73%/78.57%/81.82% precision, and 72.73%/91.67%/90% recall among Grade 1/2/3, respectively. Feature importance in MVB and tumor grading models ranged from isopropanol (-0.01) to taurine (0.04) and 3-indoxylsulfate (-0.02) to creatine (0.07), respectively. Median OS differed among Grade 1/2/3 (not reached, not reached, 181 days, respectively; $P < 3.45e-14$). This study demonstrates the use of non-invasive metabolic profiling and artificial intelligence to allow for efficient and accurate diagnosis and grading of CMTs. Future AI-powered diagnostic tools could inform human breast cancer research as well as transform lab reports and medical data into actionable insights.