Machine Learning for Voice-Based Assessment of Parkinson's Disease

Bringle, Carl (School: P A Fogelströms High School)

Abstract Parkinson's disease often affects voice and speech years before motor symptoms become clinically evident, and clinical diagnosis can take years to confirm. Machine learning applied to voice-based detection of Parkinson's disease remains an emerging field, and existing approaches have typically focused on maximized accuracy without addressing predictive uncertainty or testing performance under realistic recording conditions. Therefore, this study developed and evaluated a voice-based machine learning model for early assessment of Parkinson's disease that handles uncertainty and is robust to noise. An open Italian dataset with recordings from 24 individuals with Parkinson's disease and 22 healthy controls was used. Clinically established acoustic features were extracted from sustained vowel recordings. A logistic regression classifier was trained with participant-wise splitting. Conformal prediction was applied to allow the model to abstain under high uncertainty. Robustness was evaluated by adding simulated background noise to the recordings. On clean recordings the model achieved an ROC-AUC of 0.928 and accuracy of 0.905. AUC remained at 0.912 under added noise. With the uncertainty mechanism enabled, the model produced a definitive prediction in 68 percent of cases, and accuracy among those cases rose to 0.936. The study shows how voice-based machine learning can distinguish Parkinson's patients from healthy controls with high reliability, and the uncertainty mechanism strengthens the trustworthiness of retained predictions. These findings support voice as a promising biomarker for early assessment, but larger and more diverse datasets are needed to validate clinical use.

Awards Won:
King Abdulaziz City for Science and Technology: Economies of the Future