IOCBE 2024 Conference

The 1st International Online Conference on Bioengineering

16–18 October 2024 | Online

MDPI

Effect of Data Collection and Environment on Machine Learning Performance in Screening Dysphonia

V-BAL

IOWA Communication Sciences and Disorders

Ahmed M Yousef, Eric J Hunter

Department of Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa, USA

INTRODUCTION & AIM

Human voice acoustics are effective predictors of vocal health (Roy et al., 2013). Machine learning (ML) using voice acoustic features is a promising tool for assessing voice quality and disordered voice (dysphonia) (Idrisoglu et al., 2023). Combining publicly available datasets of normal and pathological voice recordings could enhance ML performance, as larger datasets improve its effectiveness (Hegde et al., 2019). However, the datasets represent a variety of recording conditions (e.g., equipment, environmental noise, room reverberation) which may impact ML accuracy, and the extent of this impact is unclear. This work aims to investigate how different recording conditions affect ML efficacy in screening dysphonia and classifying quality.

METHOD

Two datasets were considered. The first dataset (DS1) included voice samples from 148 individuals with voice disorders and 50 vocally-normal subjects. The second dataset (DS2), Perceptual Voice Qualities Database (PVQD), included 187 patients with voice problems and 89 without vocal issues (Walden, 2020). Subjects recorded sustained vowel /a:/ and other tasks; only the sustained vowel is analyzed. The two databases were collected under acceptable but varied conditions, including different locations, microphones, and recording equipment.

Acoustic voice features (28, including perturbation, noise, cepstral, and spectral analyses) were estimated from the recordings using MATLAB scripts and Praat software. These features were derived from each database. Seven ML models were considered: Logistic Regression (LogReg), Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting (GB), k-Nearest Neighbors (k-NN), AdaBoost (AB), and Extra Trees (ET). The ML models were trained as classifiers on acoustic features from each database and their combination. The models were evaluated to classify dysphonic and non-dysphonic voices for accuracy, Youden Index (Sensitivity + Specificity -1), and Area Under the Curve (AUC) of Receiver Operating Characteristic (ROC) curves. Model performance was compared to assess the impact of each dataset, collected under different conditions, and their combination on dysphonic voice classification.

RESULTS & DISCUSSION

ML models were trained using acoustic voice features and evaluated to classify dysphonic and normal voices. Figure 1 shows ROC curves for all models on DS1, DS2, and their combination. Performance and AUC declined for most models when datasets were merged.

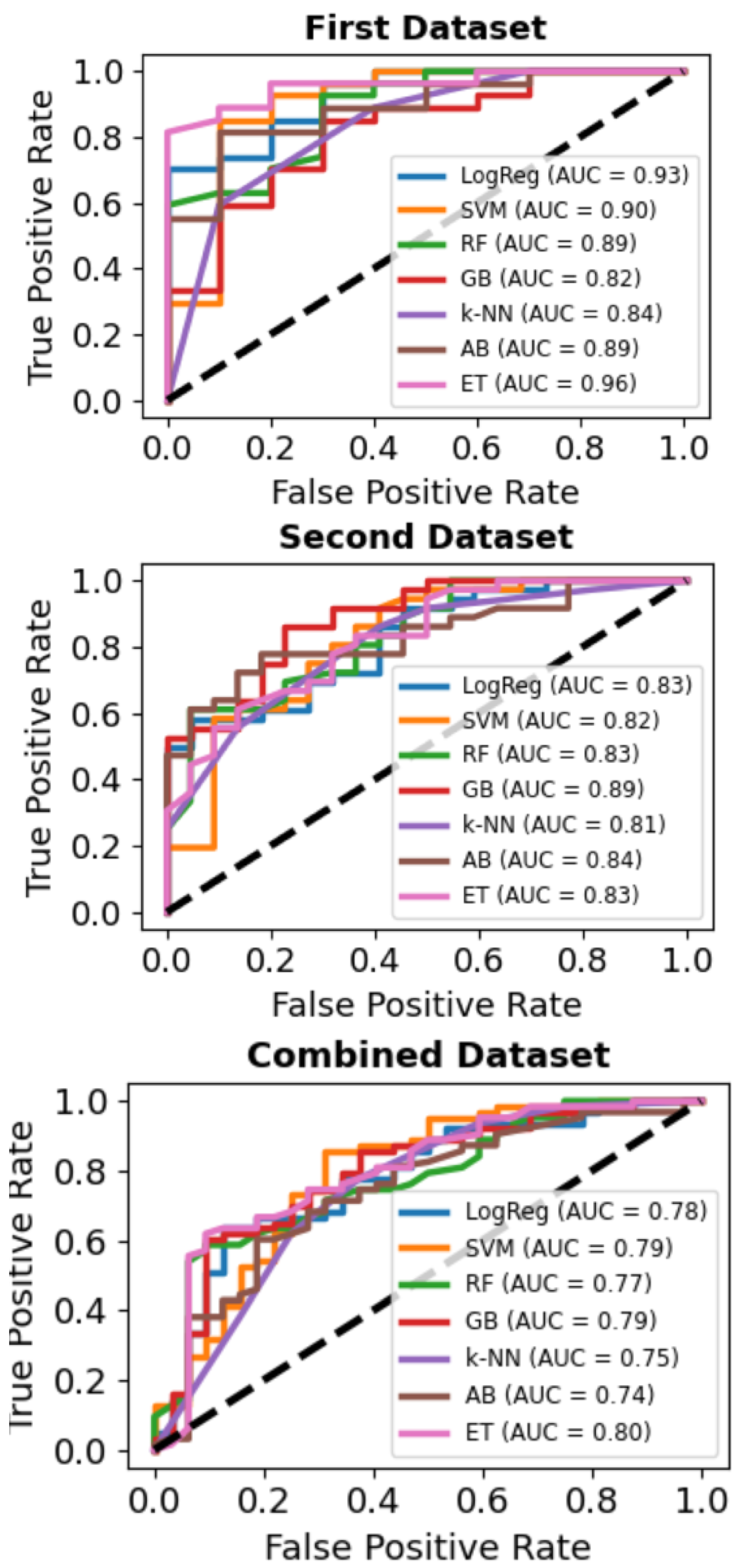

Figure 1: ROC curves and AUC values for all models on first, second, and combined dataset.

Table 1 presents two performance features and overall mean scores for models trained and evaluated on DS1, DS2, and their combination. Notably, the mean accuracy decreased from 0.79 and 0.76 to 0.73 when combining the datasets. This suggests caution when merging datasets, especially if they are collected under diverse conditions and procedures.

Table 1: Accuracy and Youden's index for all models on DS1, DS2, and both.

Model	Accuracy			Youden Index		
	DS1	DS2	Both	DS1	DS2	Both
AB	0.8	0.8	0.73	0.4	0.5	0.31
ET	0.8	0.7	0.72	0.3	0.3	0.34
GB	0.8	0.8	0.75	0.4	0.6	0.37
LogReg	0.7	0.8	0.73	0	0.5	0.36
RF	0.9	0.7	0.68	0.6	0.4	0.25
SVM	0.8	0.8	0.77	0.2	0.5	0.44
k-NN	0.8	0.8	0.73	0.4	0.5	0.4
Mean	**0.79**	**0.76**	**0.73**	**0.32**	**0.46**	**0.35**

CONCLUSION / FUTURE WORK

Combining datasets with varying collection conditions negatively impacts ML accuracy compared to using data collected under consistent conditions. Examining how factors like microphone type and room acoustics affect ML performance is critical in future to establish optimal data collection standards for ML use.

REFERENCES

Hegde, S., Shetty, S., Rai, S., & Dodderi, T. (2019). A survey on machine learning approaches for automatic detection of voice disorders. *Journal of Voice*, *33*(6), 947-e11.

Roy, N., Barkmeier-Kraemer, J., Eadie, T., Sivasankar, M. P., Mehta, D., Paul, D., & Hillman, R. (2013). Evidence-Based Clinical Voice Assessment: A Systematic Review. American Journal of Speech-Language Pathology, 22(2), 212–226.

Idrisoglu, A., Dallora, A. L., Anderberg, P., & Berglund, J. S. (2023). Applied machine learning techniques to diagnose voice-affecting conditions and disorders: Systematic literature review. Journal of Medical Internet Research, 25, e46105.

Walden, Patrick R (2020), "Perceptual Voice Qualities Database (PVQD)", Mendeley Data, v2, https://data.mendeley.com/datasets/9dz247gnyb/1