Voice for Health: The Use of Vocal Bio-markers from Research to Clinical Practice

In this paper authors discuss the new opportunities for healthcare using AI. Authors also discuss the key challenges to overcome in the near future for a substantial and efficient use of voice in healthcare. In this paper authors offer a comprehensive overview of all the present and future applications of voice for health related purposes.

voice technology is even considered as one of the most promising sectors, with healthcare being predicted to be a dominant vertical in voice applications. By 2024, the global voice market is expected to represent up to USD 5,843.8 million . In 2019, 31% of smartphone users worldwide used voice tech at least once a week [3], and 20% of queries on Google's mobile app and Android devices were voice searches.

The number of patients affected by vocal involvement diseases are very frequent. Parkinson's disease in particular where voice disorders are very frequent (as high as 89%) and where voice changes are expected to be utilized as an early diagnostic bio-marker . These voice disorders are mostly related to phonation and articulation, including pitch variations, decreased energy in the higher parts of the harmonic spectrum, and imprecise articulation of vowels and consonants, leading to decreased intelligibility. Both Alzheimer's Disease and Mild Cognitive Impairment are proven to affect the verbal fluency, reflected by the patient's hesitation to speak and slow speech rate. Voice impairment and dysarthria are frequent comorbidities in people with multiple sclerosis. Stress is an established risk factor of vocal symptoms. It was shown that smartphone-based self-assessed stress was correlated with voice features. COVID-19 is a respiratory condition, affecting breathing and voice, and causing, among other symptoms, dry cough, sore throat, excessively breathy voice, and typical breathing patterns.

To ease these vocal infected diseases researchers come up with an idea of identifying Vocal bio-marker using AI. A vocal bio-marker is a signature, a feature, or a combination of features from the audio signal of the voice that is associated with a clinical outcome and can be used to monitor patients, diagnose a condition, or grade the severity or the stages of a disease or for drug development. Identifying a vocal bio-marker is multi staged process. It consists of the following steps.

Voice Recordings: To classify the sounds emitted from a human's mouth and analyze them for disease diagnostics into 3 main categories: verbal, vowel/syllable and nonverbal vocalizations.
Data Collection Techniques: They can be grouped into 4 main categories. . Studio-based recording,Telephone-based recording,Web-based recording,Smartphone-based recording.
Audio Pre-Processing: This includes steps such as resampling, normalization, noise reduction, framing, and windowing the data. The normalization step improves the performance of feature detection by reducing the amount of different information without distorting differences in the ranges of values.
Audio Feature Extraction: There is a need to convert the audio signal into "features," meaning the most dominating and discriminating characteristics of a signal. Various methods are proposed in the literature to identify acoustic features from the temporal, frequency, cepstral, wavelet, and time-frequency domains. The correct choice of features heavily depends on the voice disorder, disease, and type of voice recording.
Audio Feature Selection and Dimensionality Reduction: Feature selection methods such as the mRMR, GramSchmidt orthogonalization allow a subset of the original feature set to be selected

without changing them. Avoid the “curse of dimensionality” dimensionality reduction methods such as PCA ,linear discriminant analysis, random forests, or stochastic neighbor embedding can be used to transform features and perform data visualization .
Training of Algorithms: Following the selection of features, machine or deep learning algorithms, such as support vector machines, hidden Markov models, CNN and RNN networks can be trained to automatically predict or classify any clinical, medical, or epidemiological outcome of interest, from vocal features alone or in combination with other health-related data. Transfer learning is another promising approach which benefits from pre-training the model on a large voice dataset.
Testing of Algorithms: In order to have reliable estimates of the performance, cross-validation and out-of-bootstrap validation techniques can be used.
Performance Metrics: Various performance metrics are used depending on specific application and the dataset including accuracy, specificity, sensitivity (recall), precision, F measure, and AUC.

Once a vocal biomarker has been identified, as with any bio-marker, the path is still long to a clinical routine use. The US Food and Drug Administration or European Medicines Agency have not approved any vocal biomarkers yet.

Ethical and Technological Challenges to Tackle: Building and sharing large databanks of highly qualified audio recordings with clinical data and identifying key vocal biomarker candidates and move from language-, accent-, age-, and culturespecific vocal biomarkers to more universal ones will be a tough challenge. Secure data collection and storage is also need to be maintain properly.