

Digital Biomarkers Development Using Multimodal AI Technology

Daniel Kim

Received April 21, 2024

Accepted July 04, 2024

Electronic access July 15, 2024

The COVID-19 pandemic has forced the world to adopt digital technology and communications everywhere in our daily lives. These developments are accelerating the tech areas such as cloud applications, collaboration services, and notably “digital healthcare” allowing healthcare providers and patients to embrace new ways of health monitoring and diagnosis via advanced technologies based on smart wearable devices, telemedicine, big data analytics, and artificial intelligence (AI) technologies. In this paper, we describe the recent progress on digital biomarkers development for health monitoring and diagnosis, especially for voice disorders. Classification of disorders is based on the machine learning (ML) models trained on multimodal data (text, voice, and bio-signals) collected from clinical study participants. Various multimodal learning/fusion techniques from data-level to decision level fusion are critically analyzed. Finally, we propose a high-level architecture for classification models and their ML modeling techniques for voice disorders by employing hybrid multimodal fusion. The hybrid fusion method is most suitable and effective since it generates features from original multimodal data at the first step, and combines those features in ensemble learning for the final classification at the second step.

Keywords: digital healthcare, digital biomarkers, telemedicine, AI, machine learning, multimodal, fusion, classification, voice disorders

Introduction

Typically, patients visit doctors, who then tap into patient records, current observations, and years of experience to make a diagnosis and craft a treatment plan. While patient records are gradually becoming digitalized, the diagnosis process is still largely manual and, on a patient-by-patient basis. Furthermore, the COVID-19 pandemic necessitated the use of digital technology for healthcare everywhere in our lives¹. The need for daily monitoring and telemedicine only increases as the population ages. The large burden on doctors, the vast amounts of patient records and data, and recent advances in AI make healthcare a field ripe for disruption.

Before, a biomarker was an objectively measured and evaluated indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. But as the world has gone digital, medicine now has access to a new type of biomarker – “digital biomarkers”^{3,4}. Digital biomarkers are consumer-generated physiological and behavioral measures collected through connected digital devices that can be used to explain, influence, and predict health-related outcomes which can vary from explaining disease to predicting drug response to influencing fitness behaviors. In essence, they are a priori collected by digital means. Therefore, the most effective classification of these measures focuses on what is being measured and the clinical insight derived from that metric.

The growth in healthcare data can be attributed to broader

access to and adoption of digital devices as well as the ability of these devices to allow doctors to continuously monitor individuals. Consumer-facing digital devices expand the population that can generate health data. Digital devices provide all internet-connected individuals with the opportunity to monitor and track their health status daily⁵. For instance, while traditional blood pressure cuff monitors provide static, infrequent measurements, wrist-worn devices can continuously record vitals over time with minimal effort from any end-user.

Smart devices, especially smartphones or smartwatches, have become the access point to the end-consumer. Lots of smartphone owners use at least one health app, many of which allow individuals to track various health measures (e.g., blood pressure, heart rate, physical activity, sleep). Many digital health technologies enable more convenient, cost-effective means to collect discrete health measures, such as blood pressure and glucose, that clinicians can draw upon in decision-making processes⁶. Additionally, the digital footprint that consumers leave when they engage with the internet, through web browsing or social media activity, provides novel data that can be leveraged for healthcare purposes.

When paired with big data analysis, this large volume of healthcare data can be leveraged to track trends and patterns for both individuals and even populations⁷. Strong correlations between health data (e.g., physical activity, heart rate) and clinical outcomes (e.g., hospitalization for cardiovascular disease) can be used to alert consumers to take precautionary steps or

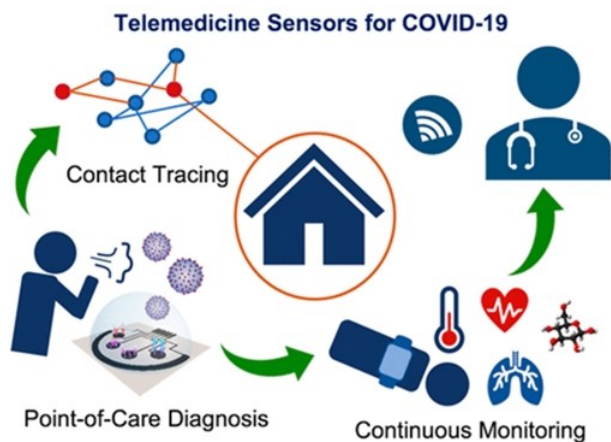


Fig. 1 . (Left) Telemedicine or mobile health during the COVID 19 pandemic to reduce potential disease spread and prevent overloading of the healthcare system through at-home COVID-19 screening, diagnosis, and monitoring¹; (Right) A demonstration of telehealth services allowing patients to communicate with doctors over audiovisual links²

present themselves to a clinical care setting for high-acuity conditions. Health systems can apply these insights to deliver the optimal treatment at the best time to each patient, including better identifying and managing high-risk, high-cost patients. For instance, a biosensor recording and analyzing intestinal vital signs is used to predict whether abdominal surgery patients will have postoperative complications.

In the realm of health determinants, researchers at prominent health organizations such as the CDC (Centers for Disease Control and Prevention) and WHO (World Health Organization) classify these factors into five distinct categories: Behavior, Social Circumstances, Environmental and Physical Influences, Genetics, and Medical Care. Notably, over 60% of health determinants fall within the realm of nonclinical factors, as highlighted by McGinnis⁸. This emphasis on nonclinical factors underscores the significance of digital biomarkers. Leveraging advanced technologies embedded in smart wearable devices and employing big data analytics, digital biomarkers offer a unique capability to measure health-related nonclinical factors while also digitizing traditional biomarkers. The amalgamation of digital biomarkers not only facilitates the measurement of nonclinical factors but also paves the way for the discovery of phenotypic signatures, enhancing our understanding of the variance in human health and disease.

Amidst the challenges in the healthcare landscape, exacerbated by the COVID-19 pandemic, the importance of AI-augmented healthcare has become increasingly apparent. Community telehealth services, particularly through videoconferencing systems, have gained prevalence during these challenging times (as seen in Figure 1), especially in the domain of mental health. However, the adoption of these services faces impediments, with health providers ill-equipped to engage effectively. Technological constraints and the social isolation conditions

imposed by the pandemic further hinder engagement. Both providers and patients grapple with the nuances of virtual interactions, impacting the assessment of non-verbal cues and overall engagement. Recognizing patient engagement as a priority is crucial, not only for optimizing retention in care but also for improving health outcomes.

Digital biomarkers find their greatest potential in supplementing existing biomarkers, particularly in medical domains reliant on subjective and observational assessments. Take neurological conditions, for instance, where disorders of the brain pose challenges for clinicians relying on empirical evaluations. Behavioral and psychosocial assessments, subject to varying sensitivity and specificity, become essential. Digital biomarkers offer a breakthrough by passively measuring digital and online behavior, providing an opportunity for more objective measures in fields where establishing gold standard metrics is elusive, such as neurology and psychiatry.

In the evolving landscape of healthcare, new solutions centered around digital biomarkers harness the capabilities of personal smart devices, advanced sensors, and cutting-edge AI and ML technologies. These solutions democratize health monitoring, allowing individuals to track their health conveniently in real-time from their homes. In contrast, traditional biomarkers, collected at specialized facilities, prove costly and inconvenient. Furthermore, the real-time availability of data from digital biomarkers provides timely insights into one's health status, revolutionizing healthcare. As the global digital biomarkers market is poised to reach \$22.54 billion by 2030, growing at an annual rate of 36.2% from 2020-2030, the transformative impact of digital biomarkers is evident, driven by the increasing demand for mobile health apps⁹.

In the subsequent sections, we first describe several notable collaboration projects for digital biomarkers discovery. In this

paper, we focus on digital biomarkers specifically for ‘voice disorders’ which will be explained later in the following section. We then describe multimodal AI digital biomarkers and further analyze various multimodal learning/fusion techniques. Finally, we propose a high-level architecture suitable for multimodal digital biomarkers for voice disorders.

Collaboration towards Digital Biomarkers Discovery

Collaboration with medical experts is key to the successful development of digital biomarkers. Tech companies are conducting digital biomarker research in cooperation with hospitals and healthcare companies.

Biogen, one of the pioneers in neuroscience, started a new virtual research study, in collaboration with Apple, to investigate the role Apple’s products, notably the Apple Watch and iPhone, could play in monitoring cognitive performance and screening for decline in cognitive health including mild cognitive impairment (MCI)¹⁰. As cognitive health—the ability to clearly think, learn, and remember—is an indicator of brain health and important to perform daily activities, the study’s primary objectives are to develop digital biomarkers to help monitor cognitive performance over time and identify early signs of MCI (an early indicator of certain forms of dementia such as Alzheimer’s disease).

Evidation Health reported the results of its exploratory research work in multiple sclerosis digital biomarkers in partnership with Novartis¹¹. The objective of the research was to explore endpoints computed from wearables that could differentiate patients with multiple sclerosis (MS) from matched non-MS patients in everyday life.

Wearable sensors may signal if a user is developing COVID-19 even if the symptoms they show are subtle. A smart ring that generates continuous temperature data may foreshadow COVID-19, even in cases when infection is not suspected. The device, which may be a better illness indicator than a thermometer, could lead to earlier isolation and testing, curbing the spread of infectious diseases, according to a preliminary study led by UC San Francisco and UC San Diego. An analysis of data from 50 people previously infected with COVID-19¹² found that data obtained from the commercially available smart ring accurately identified higher temperatures in people with symptoms of COVID-19.

Roche, headquartered in Basel, Switzerland, is a leader in research-focused healthcare with combined strengths in pharmaceuticals and diagnostics. Prothena Corporation is a late-stage clinical company with expertise in protein dysregulation and a diverse pipeline of novel investigational therapeutics with the potential to change the course of devastating neurodegenerative and rare peripheral amyloid diseases. Roche Pharma Research

and Early Development, a division of Roche Pharmaceuticals, developed a new mobile app to measure Parkinson’s disease symptoms. The app was developed in partnership with Max Little, a British mathematician at the head of the Parkinson’s Voice Initiative, and it will be used in a drug development trial with Prothena Biosciences¹³.

The app is run on Samsung Galaxy S3 Mini phones, given to participants just for the study, and stripped of most other functionality. This will help to facilitate smartphone use in the elderly patient population. Patients will do six 30-second active tests a day on the app, as well as passive monitoring. The six tests consist of a voice test (saying “aaah” for as long as possible), a balance test (standing still), a gait test (walking 20 yards and turning around), a dexterity test (tapping buttons on the touch screen), a rest tremor test (holding the smartphone and counting down from 100), and a postural tremor test (the same as the rest test, but with the hand outstretched).

Mobile health provides a unique opportunity to measure Parkinson’s symptoms more accurately and continuously than the clinical status quo, a gait test conducted during an office visit. For that reason, Parkinson’s research is also the subject of one of the initial studies being conducted via Apple’s Research Kit.

Mayo Clinic and Vocalis Health, Inc., a company pioneering AI-based vocal biomarkers for use in healthcare, collaborated to research and develop new voice-based tools for screening, detecting, and monitoring patient health. The study aimed to identify vocal biomarkers for pulmonary hypertension (PH) which could help physicians detect and treat PH in their patients. The Mayo research team and Vocalis Health first established a relationship between certain vocal characteristics and PH, subsequently conducting a prospective clinical validation study to develop PH vocal biomarkers¹⁴.

In this paper, we study voice disorders (at various levels), and advanced digital biomarkers development to classify those disorders using multimodal data from voice and bio signals.

Digital Biomarkers for Voice Disorders

Traditional ear, nose, and throat (ENT) endoscopy in hospitals involves invasive procedures, is time-consuming, and often yields non-specific results for particular diseases. On the other hand, leveraging voice biomarkers to extract relevant features through vocal recordings using mobile devices offers a non-invasive, disease-specific approach¹⁵. This advanced method utilizing digital biomarkers proves to be both time and cost-effective for patients, enhancing the efficiency of measurements. It empowers physicians to make more informed decisions, particularly in remote patient monitoring scenarios. Additionally, it encourages patients to manage their conditions more frequently and comfortably. Developing AI/ML algorithms for detecting voice diseases (e.g., Glottal Insufficiency) based on multimodal

voice/Electroglottograph (EGG) biosignals can be effectively done in collaboration with otolaryngology experts.

Voice Disorders

Voice disorders can be grouped at three different class levels - (1) normal/patient, (2) disorder groups, and (3) specific diseases. In Class Level 1, it simply distinguishes normal controls from patients. So, we have the classes of Normal, Patient (Abnormal), or Undetermined in this level. In Class Level 2, we have 8 vocal disease groups as shown below.

- BV** - Benign Vocal Mass
- MV** - Malignant Vocal Mass
- GI** - Glottal Insufficiency
- ND** - Neuromuscular Dysphonia
- ID** - Inflammatory Disease
- FD** - Functional Dysphonia
- UD** - Undetermined
- NR** - Normal

In Class Level 3, we distinguish diagnosed specific diseases such as Contact Ulcers, Cysts, Granuloma, Hemorrhage, Hyperkeratosis, Laryngitis, Leukoplakia, Nodules (nodes), Papilloma, Polyps, etc. At this level, numerous specific (not group-level) diseases would need to be accurately classified, so it is most difficult (among all the three levels) to obtain reliable accuracy of classification over more than 30 classes. To understand how these voice disorders are monitored and classified by a patient, we can think of a remote healthcare use-case such that a remote patient may use, on a daily basis, a wearable device that collect various voice and bio-signal data to be used by the digital biomarker for disorder classification (at the level of normal vs. abnormal, or at the level of specific diseases). The ENT doctor can then use this classification result to decide to bring in the patient that needs further diagnosis to the hospital.

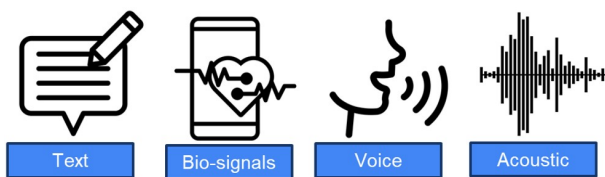


Fig. 2 Multimodal data features for voice disorders classification include texts, bio-signals, voice data and acoustic features from pre-planned clinical studies on patients for voice disorder diagnosis at hospitals while normal data samples can be collected from non-patients volunteers.

Collected Features

In order to build AI/ML classification models used in the digital biomarker software, enough training data would need to be collected from groups of patients and normal people either at hospitals or even conveniently at home. Raw sensor data are usually converted (through feature extraction) into more compact features easier to handle and also more powerful to build AI/ML (Artificial Intelligence/Deep Learning) classification models. For voice digital biomarkers, vocal features are mainly used along other metadata features as illustrated in Figure 2. Below is a typical set of those features (metadata, indices, vocal features, biosignals):

- Demographic information: Age, Gender
- Clinician-based dysphonia severity index: GRBAS
- Self-reporting clinical questionnaires:
 - VHI (Vocal Handicap Index), VFI (Vocal Fatigue Index), VAPP (Voice Activity and Participation Profile)
- Acoustic features from commercial equipment
- Voice recordings: Vowel sound (ah-), Paragraph reading
- Biosensor data: Electroglottograph (EGG) signal

While some features are simply collected by completing questionnaires or self-reports, the others would need to use specialized equipment for voice recording and bio-signal collection. It is worth noting that data collection would need pre-planned clinical studies on patients for voice disorder diagnosis at hospitals while normal data samples can be collected from non-patients' volunteers. One can leverage from the already existing data repositories from multiple different hospitals, but those data samples would require normalization for model training since their formats and/or data ranges could be different among hospitals. It is more effective to collect data in real-life and lab settings together (using data collection devices like those shown in Figure 3).

To handle the unbalanced data issue, we could apply under-sampling. We excluded ambiguous labels (i.e., normal: samples with no symptom or with ignorable symptoms; abnormal: samples with obvious dysphonia severity index). Used acoustic features such as predictors and included age, gender, age/gender interaction terms as adjusting factors. Because the acoustic features were highly correlated, to handle the multicollinearity issue and avoid overfitting the penalized logistic regression approach with a lasso penalty could be used.

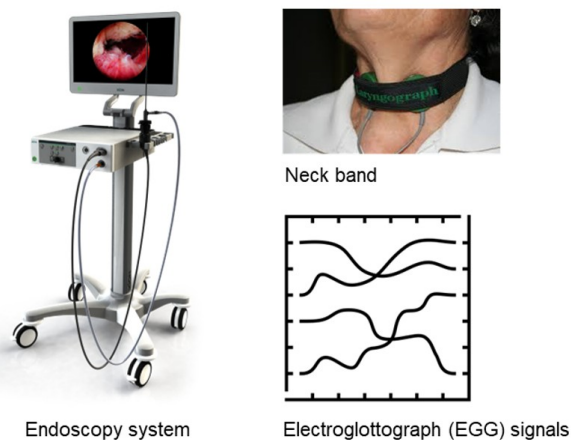


Fig. 3 (Left) Laryngoscopy is an endoscopic procedure in which a special instrument with a tiny camera is used to view the anatomy of the voice box. (Right) An example neck band that can collect Electroglottograph (EGG) signals for voice disorders diagnosis.

Multimodal AI Digital Biomarkers

Digital biomarkers for voice disorders classification can be built from three main development tasks: (1) developing a smart device to collect audio, language, and bio-marker signals; (2) creating AI model(s) to analyze said signals; (3) developing a healthcare platform to aid physicians in diagnosing and monitoring vocal chord and mental health issues.

To address the issues around AI diagnosis and telehealth, one should study AI models to assess vocal and mental health issues. Those models work best when they are based on a semi-supervised AI learning paradigm, which will jointly leverage visual, vocal, text, and bio-marker signals from patients to capture the fine-grained interplay between visual cues and language. The models will attend to specific regions in images and words in text/voice through multiple steps and gather essential information from multiple modalities to increase diagnosis efficacy and accuracy. This multimodal fused analysis will be used to derive an automated diagnosis score. From a health care perspective, an automated diagnosis score can be used in future studies to give providers real-time, quantitative feedback during or immediately post-session. Further down the line, this research can bridge the online-offline gap between physician diagnosis and self-reported surveys.

It would also require engineering to develop a device to record audio, image, and bio-marker signals. This device will help standardize received signals across patients and ensure that the research is reproducible across patients, whether receiving care in a hospital or at home.

From an AI perspective, the application of automated diagnosis for vocal cord and mental health issues is novel, and the use of a multi-modal recognition algorithm can advance the

quality of AI diagnosis research in real-world settings. From a clinical perspective, this research will help develop an approach for near real-time from patients to doctors and vice versa. Further down the line, this technology could be used to enhance physician training and help improve quality of care. The foray into tele-health will also help provide better care to under-served populations in rural areas and bridge the gap between online and offline care.

In this paper, we focus on AI digital biomarkers based on voice, language, and bio-signals. Multimodal digital biomarkers, compared to traditional self-report and equipment-based measurement, could be more effective in diagnosing and managing voice and mental disorders. Mobile devices can collect voice, text, and bio-signal data ubiquitously, which helps monitor health issues on a daily basis.

Multimodal digital biomarkers can combine digitized features related to voice and/or mental disorders, and can be built using state-of-the-art ML techniques.

In traditional AI/ML, unimodal models use data of a single type, e.g., object recognition using images or natural language understanding using voice. Multiple types of data are widely available thanks to advancement of IoT technologies and ubiquitous smart mobile devices. In modern AI/ML, multimodal data contains more information and multimodal learning is much more powerful and performs better than unimodal learning. Some examples of multimodal learning are sentiment analysis using speech (text), facial expression (vision), and voice recording (audio), stock market prediction using news articles (text), graphs of economic data (vision), and structured data.

Multimodal Fusion

In multimodal AI/ML, fusion techniques are employed to learn and decide from multimodal data at various levels from (early) data fusion to (late) decision level fusion. In this section, several multimodal fusion techniques and their applications are described. We also offer our own critical reviews on these methods to understand their characteristics and pros/cons.

Data Level Fusion (Early Fusion)

These techniques directly combine low-level data collected from multiple sources at the early stage of machine learning. They are most effective when each data type is in the same or similar form so that multiple data vectors are simply concatenated to become a large single data vector. Alternatively, correlation analysis over those multiple data sources is used to understand how they are co-related and to map the original data space to the classification label space using the mapping functions learning, such as CCA (Canonical Correlation Analysis) space mapping to nonlinear CCA. Multimodal learning using data level fusion is shown¹⁶ for multi-view recognition tasks, cross-modal retrieval/classification, and multi-view embedding.

Feature Level Fusion

Most multimodal learning methods use feature level fusion. For each modality, it performs preprocessing for feature extraction and uses fusion of feature vectors by simple concatenation in neural network models. Although a single large model combining all the features is usually smaller and more manageable than a data-level fusion model combining all the original data, it still requires too much training time and has high complexity.

Decision Level Fusion (Late Fusion)

When multimodal learning began to be studied initially, researchers typically used decision level fusion since it is relatively easy to implement. It combines the model results from each modality. It simply uses averaging or voting, and some used aggregation models and functions. Ensemble methods are also decision level fusion. These methods allow individual classifiers with rapid training (relatively) but it is hard to capture inter-relationships between modalities. One example using decision level fusion can be found¹⁷.

Hybrid Fusion

The hybrid fusion methods used in most research work for multimodal learning combine ‘learned’ (not simply extracted) features generated from deep learning (DL) models. Each DL model is trained to generate these learned features best for the final classification. You can think of this as a method of two-step DL with the first DL for feature extraction and the second DL for final classifier learning. It mitigates the disadvantages of both the data level and decision level fusion, and produces classifiers with mixtures of a variety of feature combinations and architectures. Some examples are found¹⁸⁻²⁰.

Multimodal Learning Examples

In this section, we reviewed a few multimodal learning research papers along with what modality and what fusion methods each used.

An et al.²¹ proposed a Multimodal Attention-based Fusion Networks (MAIN) for diagnosis prediction. Their multimodal fusion module based on weighted averaging is utilized to integrate the representations derived from different modalities and their correlation to obtain the patient representation for diagnosis prediction. They used the fusion of medical features and sound (vocal) features at the feature and decision levels.

Wang et al.²² described the techniques involved in each step of wearable sensor modality centred human activity recognition in terms of sensors, activities, data pre-processing, feature learning and classification, including both conventional approaches and deep learning methods. They used bio and wearable sensors for fusion at the data and feature levels.

Ariyanti et al.²³ investigated a stacked ensemble learning method to classify pathological voice disorders by combining acoustic signals and medical records. In their proposed ensemble learning framework, stacked support vector machines (SVMs) form a set of weak classifiers, and a deep neural network (DNN)

acts as a meta learner. They used medical and voice/acoustic features for fusion at the feature and decision levels. This paper is the most closely related to ours as voice disorders classification is the main focus of study.

Suggested Approach: Hybrid Multimodal Fusion

In this section, based on the literature study and analysis, we suggest an ideal AI/ML architecture for diagnostic models for vocal disorders by employing hybrid multimodal fusion. This multimodal fusion approach involves self-report questionnaires (text), medical information (text), voice signals (vectorized numerical values), and bio sensor signals (vectorized numerical values) as input.

We explain how the features for voice disorder classification are generated by using various feature learning techniques. All the textual data are put together as metadata that includes age, gender, and all the clinical indices that measure the patient’s condition (like GRBAS, VHI, VFI, and VAPP as described in the data features section). The metadata learning model uses a **random forest** technique²⁴ best known for classifying non-numerical data. Raw voice signals are converted to more compact acoustic features such as mel-frequency cepstral coefficients (MFCCs), spectral centroid, and fundamental frequency (F0) to be used in the acoustic learning model. This model uses a typical **DNN** (deep neural network²⁵ that generates ‘learned’ features best for the final classification. We suggest to use ‘spectrograms’. A spectrogram is a visual way of representing the signal strength, or “loudness”, of a signal over time at various frequencies present in a particular waveform. Since spectrograms are 2D images we use **CNN** (convolutional neural network²⁶) for learned feature generation. CNN is well known for its capability of finding inherent patterns to be used for classification, and it provides more accurate and robust classification results for 2D data (e.g., images) compared with general DNN. Lastly, raw PCM (pulse code modulation) signals are fully used to catch any hidden information that might be missed from acoustic features or spectrograms. **Transformers**²⁷ are known to be very effective when dealing with full data thanks to their attention mechanism. We suggest to use a transformer model for learning features from raw PCM signals. It should be noted that all the inputs are simultaneously used in each model since those models and their outputs are independent from each other.

Once all the ‘learned’ features from the original data of various modalities are generated as explained above, it’s suggested to use an ensemble machine learning technique called XGBoost, designed to perform final classification given a mixture of input features by properly assigning weights. XGBoost, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking prob-

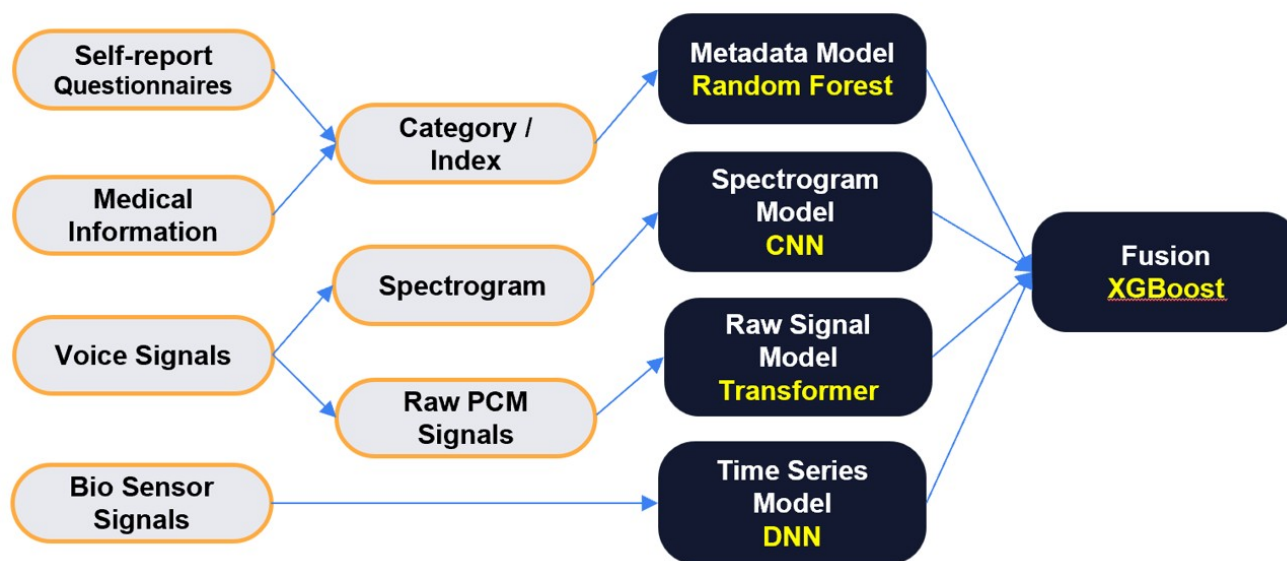


Fig. 4 A diagram of hybrid multimodal fusion of (first-step) deep feature extraction and (second-step) high-level ensemble classification from multimodal data for voice disorders classification. This serves as the main model of the digital voice biomarkers used in routine diagnosis or monitoring of patients.

lems. Those intermediate models (random forest, DNN, and CNN, Transformers) all generates predictions as output. Those output predictions are fed into the XGBoost model for final classification learning. XGBoost builds a predictive model by combining the predictions of multiple individual models in an iterative manner. The XGboost works by sequentially adding weak learners to the ensemble, with each new learner focusing on correcting the errors made by the existing ones. The hybrid fusion method used in the suggested architecture for multimodal learning combines learned features generated from first-stage deep learning (DL) models. Each DL model is trained to generate these learned features best for the final classification. You can think of this as a method of two-step DL with the first DL for feature extraction and the second DL for final classifier learning. It mitigates the disadvantages of both the data level and decision level fusion. and produces classifiers with mixtures of a variety of feature combinations and architectures. The suggested hybrid multimodal fusion shown in Figure 4.

Conclusion

We studied and described digital biomarkers development for voice disorders classification based on the ML models trained on multimodal data (text, voice, and bio-signals) collected from clinical study participants. Digital biomarkers find their greatest potential in supplementing existing biomarkers, particularly in medical domains reliant on subjective and observational assessments. Leveraging voice biomarkers to extract relevant

features through vocal recordings using mobile devices offers a non-invasive, disease-specific approach. This advanced method utilizing digital biomarkers proves to be both time and cost-effective for patients, enhancing the efficiency of measurements. Multiple types of data are widely available thanks to advancement of IoT technologies and ubiquitous smart mobile devices. In modern AI/ML, multimodal data contains more information and multimodal learning is much more powerful and performs better than unimodal learning. We suggested an ideal AI/ML architecture for diagnostic models for vocal disorders by employing hybrid multimodal fusion. It mitigates the disadvantages of both the data level and decision level fusion. and produces classifiers with mixtures of a variety of feature combinations and architectures. Digital biomarkers empower physicians to make more informed decisions, particularly in remote patient monitoring scenarios. Additionally, it encourages patients to manage their conditions more frequently and comfortably. Ultimately, it is aimed to build a remote healthcare platform equipped with discovered digital biomarkers in conjunction with AI (artificial intelligence) software that enables doctors and patients to predict and manage health problems daily remotely.

Acknowledgments

I would like to acknowledge my mentor, Dr. Mohammad Rostami, Research Assistant Professor of the Computer Science Department at the University of Southern California, for his guidance and help with AI and ML technologies.

References

- 1 H. Lukas, C. Xu, Y. Yu and W. Gao, *ACS nano*, 2020, **14**, 16180–16193.
- 2 *Healthspot Raises \$10.4M*, <https://www.bizjournals.com/columbus/news/2013/04/23/healthspot-raises-104m-from.html>, 2013.
- 3 A. Kalali, S. Richerson, E. Ouzunova, R. Westphal and B. Miller, *Handbook of Behavioral Neuroscience*, Elsevier, 2019, vol. 29, pp. 229–238.
- 4 A. Coravos, S. Khozin and K. D. Mandl, *NPJ digital medicine*, 2019, **2**, 14.
- 5 A. I. Stoumpos, F. Kitsios and M. A. Talias, *International Journal of Environmental Research and Public Health*, 2023, **20**, 3407.
- 6 A. Fabbriozio, A. Fucarino, M. Cantoia, A. De Giorgio, N. D. Garrido, E. Iuliano, V. M. Reis, M. Sausa, J. Vilaça-Alves, G. Zimatore, C. Baldari and F. Macaluso, *Healthcare (Basel)*, 2023, **11**, 1805.
- 7 K. Batko and A. Ślezak, *Journal of Big Data*, 2022, **9**, 3.
- 8 A. Spencer, B. Freda, T. McGinnis and L. Gottlieb, *Measuring social determinants of health among Medicaid beneficiaries: early state lessons*, Center for health care strategies technical report, 2016.
- 9 *Digital Biomarkers Market Report*, <https://www.strategicmarketresearch.com/market-report/digital-biomarkers-market>, 2023.
- 10 *Biogen to Launch Pioneering Study to Develop Digital Biomarkers*, <https://investors.biogen.com/news-releases/news-release-details/biogen-launch-pioneering-study-develop-digital-biomarkers>, 2020.
- 11 *Multiple Sclerosis Wearable Data Can Assess MS Patients' Disease Activity*, <https://evidation.com/resources/multiple-sclerosis-wearable-data-can-assess-ms-patients-disease-activity>, 2021.
- 12 B. L. Smarr, K. Aschbacher, S. M. Fisher, A. Chowdhary, S. Dilchert, K. Puldon, ... and A. E. Mason, *Scientific reports*, 2020, **10**, 21640.
- 13 T. Kilchenmann, B. Mollenhauer, A. Bamdadian, W. L. Popp, W. Y. Cheng, Y. P. Zhang, ... and M. Lindemann, *Scientific Reports*, 2022, **12**, 12081.
- 14 *Mayo Clinic and Vocalis Health Collaborate for Development of AI-Based Vocal Biomarker*, <https://www.healthcaretechoutlook.com/news/mayo-clinic-and-vocalis-health-collaborate-for-development-of-ai-based-vocal-biomarker-nid-2059.html>, 2021.
- 15 J. D. S. Sara, D. Orbelo, E. Maor, L. O. Lerman and A. Lerman, *Mayo Clinic Proceedings*, 2023, **98**, 1353–1375.
- 16 C. Guo and D. Wu, *arXiv preprint arXiv:1907.01693*, 2019.
- 17 S. C. Huang, A. Pareek, R. Zamanian, I. Banerjee and M. P. Lungren, *Scientific reports*, 2020, **10**, 22147.
- 18 M. R. Amer, T. Shields, B. Siddiquie, A. Tamrakar, A. Divakaran and S. Chai, *International Journal of Computer Vision*, 2018, **126**, 440–456.
- 19 H. Xu, W. Liu, J. Liu, M. Li, Y. Feng, Y. Peng, ... and M. Wang, *Proceedings of the 3rd International on Multimodal Sentiment Analysis Workshop and Challenge*, 2022, pp. 15–21.
- 20 Y. Wang, J. Peng, J. Zhang, R. Yi, Y. Wang and C. Wang, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 8032–8041.
- 21 Y. An, H. Zhang, Y. Sheng, J. Wang and X. Chen, *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, pp. 809–816.
- 22 Y. Wang, S. Cang and H. Yu, *Expert Systems with Applications*, 2019, **137**, 167–190.
- 23 W. Ariyanti, T. Hussain, J. C. Wang, C. T. Wang, S. H. Fang and Y. Tsao, *IEEE Sensors Letters*, 2021, **5**, 1–4.
- 24 G. Biau and E. Scornet, *Test*, 2016, **25**, 197–227.
- 25 V. Sze, Y. H. Chen, T. J. Yang and J. S. Emer, *Proceedings of the IEEE*, 2017, **105**, 2295–2329.
- 26 S. Albawi, T. A. Mohammed and S. Al-Zawi, *2017 international conference on engineering and technology (ICET)*, 2017, pp. 1–6.
- 27 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, *Advances in neural information processing systems*, 2017.