VoiceEDx: A Novel, Voice-Based End-to-End Multi-Disease Diagnostic Platform Using a Highly Accurate, Expandable Artificial Intelligence Engine for an Early, Secure and Reliable Diagnosis of Disease

Project Overview

The current problem in the field of disease diagnostics is the lack of an accurate, accessible, and low-cost diagnostic method capable of early detection. Consequently, millions of people across the globe suffer from misdiagnosis and the lack of a diagnosis altogether, leading to ineffective treatment plans, decreased quality of life, and in many cases, premature death. In this project, we combat this problem by creating the first end-to-end software solution to diagnose 18 different diseases through voice dysphonia - a novel biomarker for disease diagnostics.

Voice Dysphonia, one of the earliest observable symptoms in hundreds of different conditions, can lead to a proper diagnosis if detected early. In this research, a user-submitted eleven (11) second voice sample was used as the sole medium of diagnosis, with a custom-built machine learning engine generating a 95%+ accurate diagnosis for all disease supported. Additionally, the platform is highly expandable, as diseases are seamlessly integrated into the platform. Overall, this project represents a major shift in the field of disease diagnostics, and is the only diagnostic tool to-date that fully diagnosis a plethora of different diseases in a non-invasive, low-cost manner.

Background

Disease	Existing Diagnostic Accuracy	Voice Dysphonia
Lung Cancer	N/A	Yes
Cardiovascular disease	90%*	Yes
Concussion	46%	Yes
Traumatic Brain Injury	65%	Yes
Autism	N/A	Yes
Dementia	79.8%	Yes
Parkinson's + Essential Tremor	26%	Yes
Alzheimer's	46%	Yes
ALS	48%	Yes
Huntington's	62%	Yes
Laryngeal Cancer	77%	Yes
Hypopharyngeal Cancer	68%	Yes
Multiple System Atrophy	56%	Yes
Multiple Sclerosis	56%	Yes
Cerebral Palsy	65.5%	Yes
PTSD	63%	Yes
ADHD	N/A	Yes
Depression	33.3%	Yes

In our research, we found and implemented a new biomarker for disease diagnostics - voice. Voice Dysphonia was observed in 18 unique diseases and used to yield a highly accurate, early, and reliable diagnosis of disease.

The diagnostic procedure for all 18 of these diseases involve some form of doctor-patient interaction, meaning that accessible diagnosis is limited. In this project, we aimed to alleviate this problem while also drastically improving diagnostic accuracies with a focus on early detection. Voice, the earliest observable symptom in many different diseases, was found to be a suitable biomarker after thorough analysis.

Affected vs. Healthy Sample

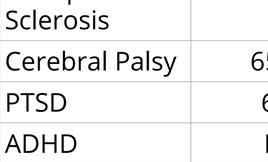
The machine learning platform is trained to detect minute variations in the normal voice pattern that distinguish users with the disease from those without the disease. These variations typically relate to their unsteady fundamental frequency but they also can relate to the hoarseness in their voice or lack of control over their vocal cords. There are also mild variations in other features depending on affected area. The differences in voice patterns are shown below.

Healthy

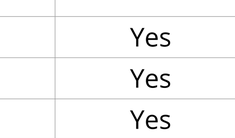
- In comparison to those with Parkinson's disease, they have relatively steady voices
- Low variation in fundamental frequency along with more straight pitch contours

Disease

- Typically have unsteady or hoarse voices
- Higher variations in fundamental frequency along with a jagged pitch contour



```
0.0026600000
0.0000221970
0.0014300000
0.0016800000
0.0042800000
0.0274400000
0.2400000000
0.0147100000
0.0180200000
0.0214400000
0.0441300000
0.0159370000
18.4800000000
```



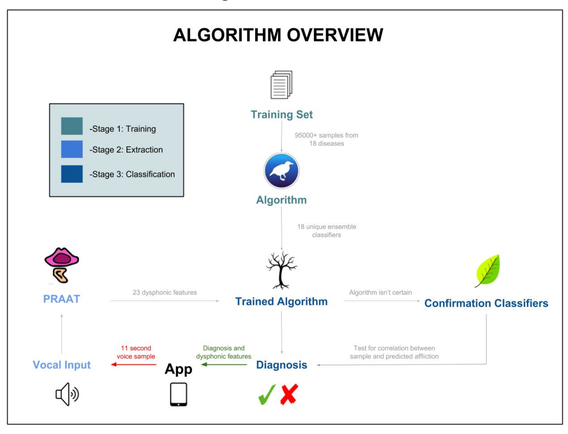
```
0.0059600000
0.0000388060
0.0030900000
0.0034200000
0.0092800000
0.0365600000
0.3150000000
0.0166900000
0.0246200000
0.0405800000
0.0500800000
0.0182250000
18.9940000000
```

Voice patterns in Healthy vs Parkinson's disease

Algorithm Overview

1. **Training:** The first step is to train each of the algorithms from a dataset of over 95,000 samples (see dataset for specifics on variation). Each algorithm is a variation of a heterogeneous ensemble via stacking (see section multi-disease classification for specifics on variations for each general model).

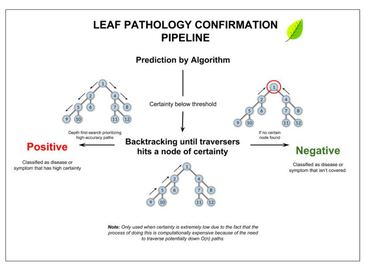
2. **Extraction:** Once the algorithm receives the 11-second vocalization of /a/ from the user, 23 dysphonic features are extracted via PRAAT (see Feature Extraction for complete listing of features and sub-stages).



3. **Classification:** Once the algorithm is trained and the features are extracted, they are transferred to classification. The algorithm uses a tree-based binary decomposition architecture to simplify into a series of binary classification problems (see multi-disease classification for more details).

Border Classification

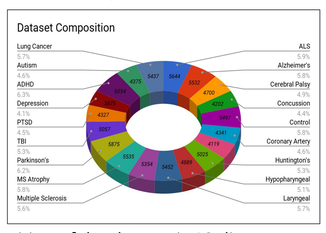
- An algorithm is trained with the positive cases being the disease being tested for and the negative cases being randomly picked samples of other diseases intended to emulate general noise.
- If the algorithm concludes that it strongly correlates with the disease in question then the algorithm returns a positive.
- If the algorithm tests it as negative or is sufficiently uncertain about the result then it is sent to a border classification system.



- The system traverses up the tree to the last certain node. Then it DFS traverses down the other sections, higher certainty priority, in order to find any diseases that might indicate a high accuracy.

Dataset

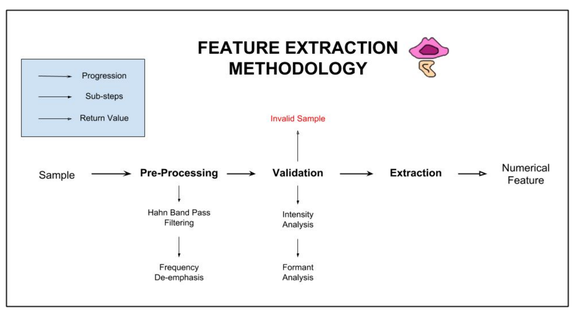
- The dataset is a selection of 95,270 data samples from 18 different diseases plus a control set. These disease datasets have been collected from a variety of different sources and all have been tested for a strong correlation between dysphonia and pathology.
- Since the majority of sources were in disparate forms, extensive data scrubbing and cleaning operations were made in order to standardize the dataset and to eliminate outliers. After that, instance selection algorithms (i.e. SMOTE and Stratified Remove Folds) were used.



The composition of the dataset is 18 diseases + control dataset with an average ~5000 samples per dataset.

Feature Extraction

In feature extraction the algorithm translates vocal data into numerical data that can be readily analyzed by the algorithm. In order to perform the feature extraction, a series of macro script was designed. There are 3 main steps of feature extraction: Pre-Processing → Validation → Extraction.



Step 1: Pre-Processing

The first stage is to filter out any extraneous noise and attempt to isolate the patient's voice. Our approach compensates for insufficient mic quality and unsuitable samples by using two algorithms: Hahn Band Pass filtering (to remove noise that is recurring at a standard rate) and Frequency De-emphasis (to isolate any frequencies that are outside the range of voice).

Step 2: Validation

The second step is to validate the quality of the sample. If the recording isn't appropriate for further analysis, the system is able to filter adequate samples from poor ones using format and intensity analysis as a validation mechanism. We use the intensity analysis to ensure that the user maintains a steady phonation and use formant analysis to ensure that the user is correctly saying /a/.

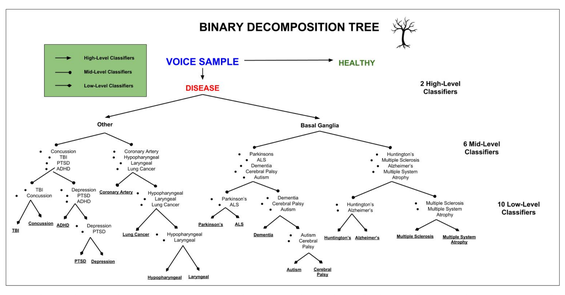
Step 3: Extraction

The last step is extracting the features from the algorithm. Here, we extract **23** dysphonic features (see right) from the post - processed sample via zero-point methodology. These features were selected by aggregating a multitude of feature selection algorithms and reviewing existing literature on acoustical analysis.

Jitter	Shimmer	MFCC	Noise
Jitter:(%)	*Shimmer: APQ*	*11th Coef*	*HNR*
Jitter: DDP	*Shimmer: (dB)*	*12th Coef*	*NHR*
Jitter(Abs)	*Shimmer: APQ 3*	*8th Coef*	*GNE*
Jitter: PPQ	*Shimmer: APQ 5*	*5th Delta - Delta Coef*	*GQ*
Jitter: RAP	*Shimmer: DDA*	*4th Delta Coef*	*VFER*
	Shimmer	*9th Coef*	
		7th Coef	

Multi-Disease Classification

In order to decompose the multi-class problem into a series of binary classifiers, the algorithm uses a tree-based method. Instead of using a brute force O(n) design architecture, the system leverages the inherent similarities of the diseases in order have an ordered binary tree which reduces it down to an O(log n) traversal solution.



The algorithm is set up to have 3 levels of classifiers plus a set of border classifiers and multi-stage classifiers.

- The schema is set up in order to be nested in groups based on their affiliation. This is based on the fact that diseases that have similar effects on the body have similar symptoms which is reflected in similar patterns.
- We leveraged this property in order to implement the multi-class composition into a binary tree format which allowed for reduced runtimes.
- This also side-steps the problem of class imbalance, that lead to problems such as local minima, which would be encountered in a OVA multi-class decomposition system.

Ensemble Learning

At each of the nodes we placed a heterogeneous ensemble method. In lieu of using a "wisdom of the crowd" methodology we decided to use a "panel of experts" approach to ensemble design. This calls for us using a series of strong classifiers that have different base approach philosophies. We then combined them via stacking with an SVM meta-classifier in order to leverage decorrelations and correct errors.

Classifiers

1. **Regression** - robust to noise and overfitting
   a. AdaBoost M1 Weakened Support Vector Machines (SVM)
   b. Primal Estimated sub-GrAdient SOlver for SVM (Pegasos SVM)
2. **Tree-Based** - well adapted to non-linear features and large datasets
   a. Random Forest
   b. Boosted REP Tree (Gradient Boosted Tree)
   c. CART
3. **Nearest-Neighbors** - extremely robust to noise
   a. Locally-Weighted Learning (LWL)
   b. k-Nearest Neighbors (kNN)
   c. K-Star
4. **Bayesian Classifier** - simple model and quick convergence
   a. BayesNet
5. **Rule Learner** - easily interpretable and highly adaptable
   a. Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
   b. MODLEM
   c. Fuzzy Unordered Rule Induction Algorithm (FURIA)
   d. Multi-Objective Evolutionary
6. **Gaussian Classifier** - high flexibility and simplicity
   a. Radial Basis Function (RBF) Classifier
7. **Misc** - chosen mainly for unique approach and high accuracy
   a. CHIRP
   b. Fuzzy Lattice Reasoning (FLR)

Multi-Stage Classification

Integrated into the algorithm is the ability to detect stages based on the severity of the symptoms and other underlying progression based factors that are unique indicators of the stage of pathology. These predictors are extensions of our normal tree based pipeline situated at the leaf nodes after the border classification pipeline. These diseases were chosen to have stage detection because of the presence of underlying progression-based milestones which act to allow the algorithm to make accurate stage diagnosis.

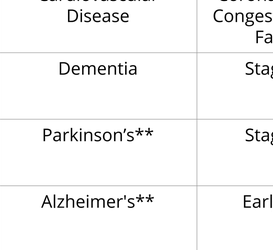
Disease	Stages	Accuracy
Lung Cancer**	Non-Small Cell: 1-4 Small Cell: 1	**93.2% - 94.6%** **92.9% - 94.3%**
Cardiovascular Disease	Coronary Artery Congestive Heart Failure	**97.66%** **94.3%**
Dementia	Stage 3-7	**91.7% - 94.37%**
Parkinson's**	Stage 1-5	**92.23% - 94.16%**
Alzheimer's**	Early-Late	**93.43% - 94.73%**
See Full List on Display Table	--	--

**Note that these accuracies are representative of a multi-class 10-fold cross-evaluation test of the discriminative capabilities of the the leaf node classifiers with only test cases of the disease in question.*
***All stages of the disease are detected*

Implementation

Throughout the development of this project, we placed an emphasis on ease-of-use. One of the most important aspects of this project is accessibility along with the scalability accomplished by the mobile app. As many people around the globe do not have access to proper medical facilities, one of the main goals of this project was to provide a universal platform for the tool.

*Main Menu of the **VoiceEDx Web Application** This website is designed to provide the full functionality of VoiceEDx online*



Results (left) and Main Menu (right)

The application also offers medical receipts, a way of exporting a VoiceEDx diagnosis to medical professionals. Receipts include a detailed analysis of a patient's voice as well as diagnostic afflictions for several medical conditions. These receipts can be used and seen by physicians, specialists, and speech therapists alike.

Backend & Blockchain

With the variety of solutions VoiceEDx offers, a unique backend approach utilizing blockchain technologies was constructed to ensure all workers operate in harmony. All processing and computational work occurs on the server-side on our custom AWS EC2 instance. The suite of mobile applications, web application, and tele-diagnosis system all utilize the HTTPS protocol to communicate with VoiceEDx servers. The backend, using *express* for middleware connections, listens to requests at certain endpoints. This allows for third-party medical facilities and developers alike to take advantage of the VoiceEDx Artificial-Intelligence (AI) engine through our REST API.

In addition to using Firebase technologies for user - management and user-demographics, VoiceEDx securely, and efficiently stores diagnostic data including afflictions and stage-classifications. The Node.js backend oversees the creation of private cryptography keys under the ECDSA standard. The manner in which these "transactions" are registered on the blockchain network, makes linking a diagnosis to a user virtually impossible. The decentralized network and usage of blockchain practices truly puts patients in control of their own data. Aside from the handling of data, the VoiceEDx backend was developed with scalability in mind. Our infrastructure is capable of serving millions of users at one moment in time. Utilizing AWS Elastic Beanstalk load-balancing technologies, as server traffic increases, duplicate instances of the backend will be launched. Our purpose in building this type of extensive infrastructure is to take the project away from a theoretical approach and to implement the system into an end-to-end solution.

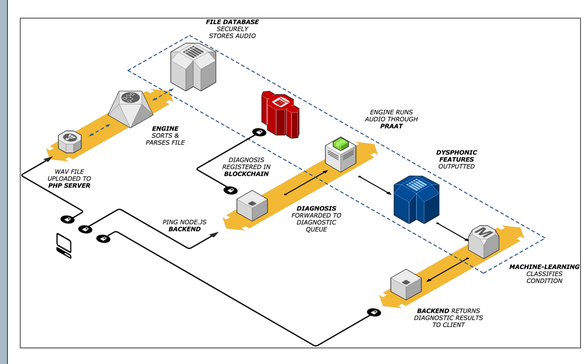


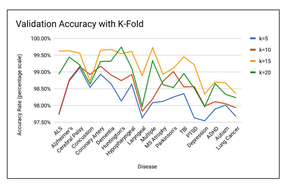
Diagram of System Architecture

Results

Qualitative Evaluation

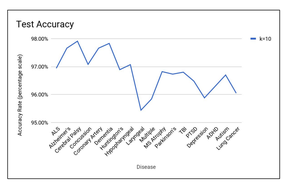
Using Multi-Stage classification, all 18 disease were detected before current clinical diagnostic methods. Accuracies for all 18 diseases were over 97%, with an average accuracy of 97.6%. Overall, the diagnostic processing takes roughly 15 seconds from start to finish to complete. In addition, diagnosis is completely automated and significantly outperforms existing methods (see *Background*).

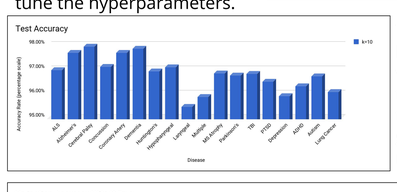
Statistical Evaluation



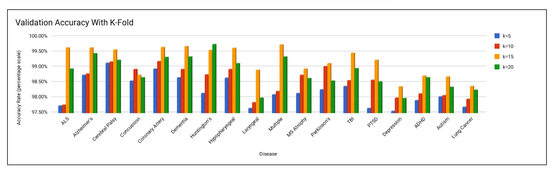
The k-fold accuracy varies depending on the amount of folds on the data, however nearly all tests have over 98% accuracy with some notable variations: PTSD, Depression , Lung Cancer, and Laryngeal Cancer.

For the test set we used a 25%-75% split with 25% being the test set. The samples randomly chosen from each set are proportional to the number of samples in each. They were used purely for final testing and not used to tune the hyperparameters.

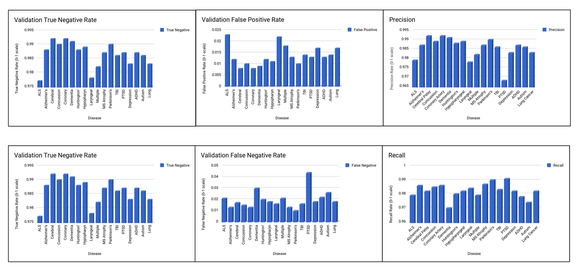




The close fit between the validation and test indicates that the algorithm is generalizable and not overfitting.



Due to the design of K-fold cross evaluation the accuracy flows in an approximate bell curve with k=15 being the apex. However, the results are usually heavily correlated.



The false positive and false negative rates were relatively equal indicating that the algorithm has no apparent bias towards any disease.

Expansions

Additional developments can be made in order to further advance and expand the project. To improve the accuracy of the classification algorithm, the hyperparameters must further be optimized. In addition, the pathfinding should also be improved for the border classification algorithm, particularly the reassignment protocol. More feature extraction algorithms could also be implemented to reduce margin of error and improve accuracy of the platform. The addition of new features and dimensionality reduction to remove redundant features can also be an area of focus in the future.

With the implementation of additional datasets and new target diseases, the reach of the project will undoubtedly improve. The quality of the existing datasets, with the utilization of more advanced instant selection algorithms and more comprehensive data scrubbing, can be improved as well. In low-resource environments, particularly those in third-world countries, improved telediagnosis and monitoring, where users no longer need a smartphone to receive a diagnosis, is an area of future work for this project.

*All images made by finalists