VocalisHealth

VocalisCheck - COVID-19

VocalisCheck Biomarker for COVID-19: A Novel Biomarker from Vocalis Health

Abstract

The COVID-19 pandemic continues to provide major challenges for governments, businesses, healthcare systems and people around the world. In addition to managing the health of people who are infected and maintaining protocols to limit the virus' spread, all of these institutions are seeking ways to safely plan for a return to work / healthcare / travel / leisure. This is especially challenging with a disease that is highly infectious and largely asymptomatic, where testing is burdensome with limited availability, treatments are as yet unproven, and vaccines are not forthcoming in the near term.

Safe business continuity planning calls for a readily accessible first-line screening tool that can reliably estimate risk without the burden / limitations of current detection methods and with more sensitivity than fever monitoring. Current tests typically have rigorous laboratory specifications, require time before results are available and are expensive. On the other hand, utilizing body temperature screening is far less accurate, with the majority of COVID-19 positive patients exhibiting no fever, and with the potential for people to mask fever with common over-the-counter medications.

In an effort to address these needs, Vocalis Health developed a Vocal Biomarker candidate to screen COVID-19. VocalisCheck is a software solution that screens for the risk of COVID-19 by analyzing a person's voice, without additional information or prior knowledge about the individual user. A simple voice sample is collected via a mobile app and/or a website browser, and a cloud-based analytics engine provides the risk score within seconds. Preliminary and promising results outlined in this White Paper support the use of the Vocalis Health Vocal Biomarker as a screening tool for COVID-19 risk, demonstrating high negative predictive value (99.2%) and higher sensitivity than fever (54.5% compared to 9.1% for fever).

The VocalisCheck COVID-19 Vocal Biomarker represents an accessible, cost-effective screening tool that can be reliably conducted at home before entering public places in a safe and efficient manner.

Background

On March 11, 2020, the World Health Organization declared the novel coronavirus disease (COVID-19) outbreak a pandemic. The disease was first reported in late December 2019 in Wuhan Province, China, and has since spread to more than 190 countries and territories globally. As of May 22, 2020, 5,268,694 confirmed cases and 338,104 deaths have been reported worldwide, and both numbers continue rising.[1] In response to the steep spread trajectory of COVID-19, governments across the world have implemented aggressive isolation strategies which slowed the transmission, but which also caused major disruptive changes to everyday life and triggered an economic recession.

The COVID-19 infection causes clusters of respiratory illness and is associated with ICU admission and high mortality rates, especially in older populations or people with underlying illness. Common symptoms include fever, cough, shortness of breath and myalgia or fatigue, but symptoms vary dramatically between patients and the vast majority of infected patients remain asymptomatic.[2,3,4] Two studies summarizing data of 1,099 and 5,700 admitted patients with laboratory confirmed COVID-19 in China and the New York City metro area reported that only 43.8% and 30.7% of patients had fever on admission, respectively.[5,6]

Current diagnostic testing methodologies such as polymerase chain reaction (PCR)–based methods or deep sequencing play an indispensable role in identifying infected patients and helping to prevent the virus from spreading. However, nucleic acid testing has rigorous laboratory specifications and requires a long time before results are available. In addition, these methods rely heavily on the presence of the viral genome in sufficient amounts at the site of sample collection, and therefore can provide false-negative results at early stages of the disease.[7,8]

Body temperature screening (fever) is the commonly used test performed to screen for risk at points of entry (e.g., in healthcare settings, factories, airports, retail). However, recent reports on asymptomatic contact

[1] The World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report – 97. 2020. Available from:https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200426-sitrep-97-COVID 19.pdf?sfvrsn=d1c3e800_6
[2] Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020.
[3] Yu P, Zhu J, Zhang Z, Han Y, Huang L. A familial cluster of infection associated with the 2019 novel coronavirus indicating potential person-to-person transmission during the incubation period. J Infect Dis 2020.
[4] Bai Y, Yao L, Wei T, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA 2020;
[5] GUAN, Wei-jie, et al. Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine, 2020.
[6] RICHARDSON, Safiya, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA.
[7] GUO, Li, et al. Profiling early humoral response to diagnose novel coronavirus disease (COVID-19). Clinical Infectious Diseases, 2020.
[8] LI, Yan; XIA, Liming. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. American Journal of Roentgenology, 2020, 1-7.

transmission of COVID-19 and false-negative results of the symptoms-based screening challenge this approach as fever screening may miss individuals incubating the disease[9].

Given the public health, economic and humanitarian implications of the COVID-19 outbreak, there is a strong need for easily accessible detection and monitoring methods that are superior to fever. Deploying such solutions will help enable safer business continuity plans, slowing transmission when people return to work / travel / leisure / healthcare and helping to reduce the likelihood of future outbreaks.

The Promising Potential of Voice in Detecting the Risk of COVID-19 Infection

Voice is a non-invasive, passive signal which can serve as a biomarker to screen and monitor health. Voice analysis has been linked to a number of health disorders including Parkinson's Disease, Obstructive Sleep Apnea and Autism Spectrum Disorder.[10,11,12] Vocalis Health is a leading pioneer in the development of vocal biomarkers for risk screening and disease monitoring. The company recently published results demonstrating the prediction of adverse outcomes (such as hospital readmission and mortality) among chronic patients.[13,14] In addition, Vocalis Health developed a first-of-its-kind tool for the detection of changes in shortness of breath in patients diagnosed with Chronic Obstructive Pulmonary Disease (COPD). As the company realized the critical role its technology might serve in the battle against the COVID-19 pandemic, Vocalis Health launched several research partnerships to reliably and quickly collect data, in order to develop a unique biomarker to assess the risk that people may be COVID-19 infected.

Data Collection

Data was collected from 1,616 participants using three methods:

- ***Clinical trial:*** A prospective, multi-center, observational clinical study which included patients with a positive COVID-19 test and medical staff members with a negative COVID-19 test. Centers

[9] BWIRE, George M.; PAULO, Linda S. Coronavirus disease-2019: is fever an adequate screening for the returning travelers? Tropical medicine and health, 2020, 48.1: 1-3.
[10] Bonneh, Yoram S., et al. "Abnormal speech spectrum and increased pitch variability in young autistic children." Frontiers in human neuroscience 4, 2011; 237.
[11] Uma Rani K., Holi M.S. "Automatic detection of neurological disordered voices using mel cepstral coefficients and neural networks". In: 2013; IEEE Point-of-Care Healthcare Technologies (PHT):76-79. Bangalore, India: January 16-18, 2013.
[12] Goldshtein, Evgenia, Ariel Tarasiuk, and Yaniv Zigel. "Automatic detection of obstructive sleep apnea using speech signals." IEEE Transactions on biomedical engineering 58.5 (2010): 1373-1382.
[13] Maor, Elad, et al. " Vocal Biomarker is Associated with Hospitalization and Mortality among Heart Failure Patients." Submitted to the Journal of American Heart Association, 2019.
[14] Sara, Jaskanwal Deep Singh et al. "Non-invasive vocal biomarker is associated with pulmonary hypertension." PloS one vol. 15,4 e0231441. 16 Apr. 2020, doi:10.1371/journal.pone.0231441.

included the Sheba Tel-Hashomer hospital (n=9), Rabin Medical Center (n=6) and IDF (Israeli Army) personnel (n=34) in Israel.

- ***Online survey:*** Large-scale, crowdsourced data collection outreach which included an open call online for anyone (healthy or diagnosed with COVID-19) to donate their voice and join the study (n=1,241).
- ***YouTube audio:*** An active search online for interviews featuring individuals diagnosed with COVID-19 and healthy individuals on YouTube (n=326).

Participants who took part in the clinical trial signed an informed consent form and their demographic and medical data was documented by research coordinators. Participants who joined the online survey self-reported their demographic and medical data using the research application. All participants recorded their voice and completed a symptom questionnaire using either a smartphone, personal computer or tablet.

Data Pre-Processing and Analysis

Since the data included recordings from an online survey, quality tests were performed to ensure these recordings suited the instructions and were of high quality. Trained personnel listened to the online survey recordings and removed recordings if they were noisy, low quality, contained less than 10 seconds of speech, or if participants did not follow the recording instructions. During this procedure, 399 recordings were excluded from the original 1,241 online survey recordings, resulting in 842 high-quality recordings from the online survey and a total of 1,217 recordings from all three methods of data collection (including one recording per participant).

All recordings were sampled at a frequency of 44.1 kHz and normalized between a range of (−1) and 1. The first 10 seconds of continuous speech in each recording was used in the analysis.

For the purpose of different analyses described in this paper, the data were divided to form a training set and a test set several times, resulting in different sets for each analysis. In each analysis, the same feature extraction and model evaluation processes were conducted on the training set, and a validation procedure was performed on the mutually exclusive test set. The results of the biomarker performance are described in detail at the end of each analysis.

The Feature Extraction Process

The feature extraction process was based on transfer learning and adaptation methods (similar to Kumar[15]). All recordings were down-sampled to 16 kHz and a spectrogram was computed using the Short-Time Fourier Transform. Each spectrogram was passed to a pretrained deep convolutional neural network (CNN), which resulted in a 512-dimensional-features vector for each recording (figure 1). This approach allows one to reach state-of-the-art results on small training databases.

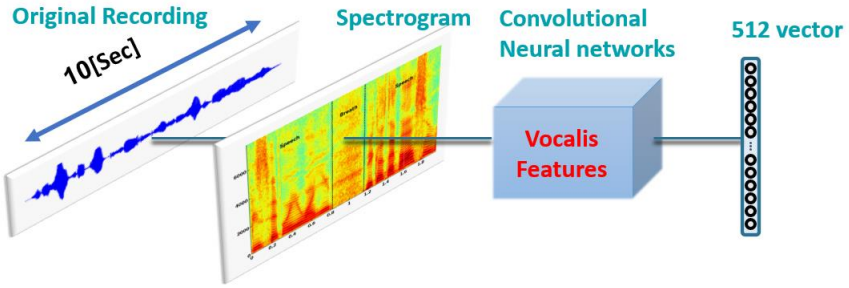

Figure 1. Feature extraction process conducted in each analysis using transfer learning and adaptation methods. The process produces a 512-dimensional-features vector for each 10-second recording.

Biomarker Training and Model Evaluation Process

A 10-fold cross-validation (CV) procedure was performed and various models were evaluated (k-nearest neighbours, support vector machine, and random forest) at different regularization levels. In each analysis, the results of the models were evaluated by the average area under the ROC curve (AUC). The chosen model in each analysis is described as the vocal biomarker, a positive scalar between 0-1, which is a non-linear combination of the 512 features mentioned above.

Analysis 1: Assessing the Vocalis Biomarker Performance on a Balanced Dataset

Our data (n=1,217) included 6.4% of participants who were COVID-19 positive (n=78) and 93.6% COVID-19 negative participants (n=1,139). Accordingly, we first constructed a balanced training set in which each positive participant was paired with a negative participant who spoke the same language, had the same gender and was a similar age (no more than a 1-year difference). This training set included a total of 156 participants from all three methods of data collection. It is important to note that some of the recordings contained scripted speech (clinical trial, online survey) while others contained free speech (YouTube).

[15] Kumar, Anurag, Maksim Khadkevich, and Christian Fügen. "Knowledge transfer from weakly labeled audio using convolutional neural network for sound events and scenes." *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018.

Baseline characteristics of the balanced training set are summarized in Table 1.

Table 1 - Baseline characteristics of the balanced training set

	Total participants (n=156)	Positive COVID-19 (N=78)	Negative COVID-19 (n=78)
Age (y)	35±14	35.6±14.2	34.34±14
Males	108 (69.2%)	54 (69.2%)	54 (69.2%)
Language:			
Hebrew	64 (41.0%)	32 (41.0%)	32 (41.0%)
English	92 (59.0%)	46 (59.0%)	46 (59.0%)
Study method:			
Clinical trial	49 (31.4%)	32 (41.0%)	17 (21.8%)
Other	107 (68.6%)	46 (59.0%)	61 (78.2%)

The feature-extraction and clinical-validation (CV) procedures described above were performed on each recording in the balanced training set. The optimal result of our 10-fold procedure was an AUC of 0.69, using a support vector machine model with a nonlinear kernel (radial basis function), as displayed in the ROC curve in Figure 2. To further validate the CV process, we randomized the labels (positive/negative) between the recordings and received an AUC 0.5-0.53, which is equal to a random classifier. This test validated there was no data leakage between the folds.

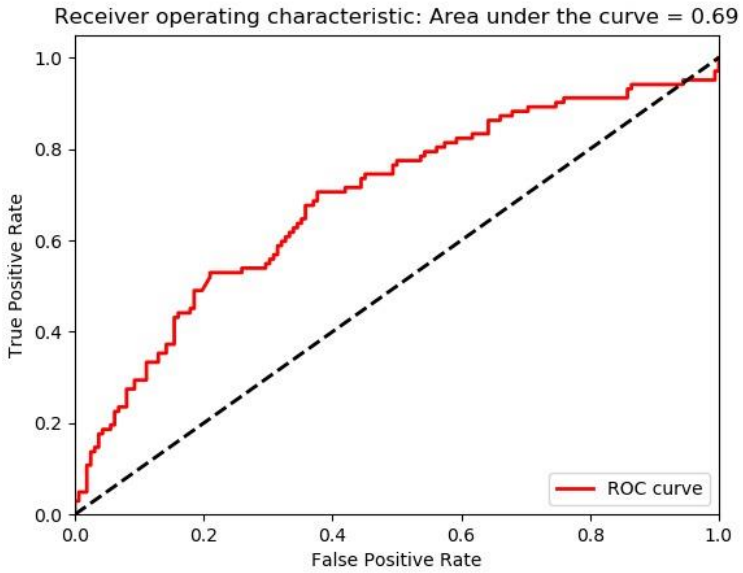

Figure 2. The average area under the curve (AUC) for COVID-19 detection (positive/negative) in the balanced training set (AUC=0.69).

Analysis 2: Assessing the Vocalis Biomarker Performance on a Free-Speech Dataset

In order to evaluate the capability of the biomarker to operate using free speech, we created a new training set with the exclusion of the YouTube recordings (27 positive) from the training set described in the first analysis. The new training set included a total of 129 recordings (a single recording per participant) which contained only scripted speech (either counting or reciting a predefined phrase). The characteristics of the new training set can be seen in Table 2.

Table 2 - Baseline characteristics of the scripted speech training set

	Total participants (n=129)	Positive COVID-19 (N=51)	Negative COVID-19 (n=78)
Age (y)	33.8±13.7	33±13	34.34±14
Males	93 (72.1%)	39 (76.5%)	54 (69.2%)
Language: Hebrew English	 64 (49.6%) 65 (50.4%)	 32 (62.7%) 19 (37.3%)	 32 (41.0%) 46 (59.0%)
Study method: Clinical trial Other	 49 (38.0%) 80 (62.0%)	 32 (62.7%) 19 (37.3%)	 17 (21.8%) 61 (78.2%)

The free speech test-set was comprised of 326 YouTube audio clips of 326 individuals (a single clip per individual), all containing free speech. Twenty-seven of the individuals were self-described as positive for COVID-19 and 299 were COVID-19 negative. The clips of the negative group were recorded prior to the end of 2018, approximately one year before the first reports of COVID-19. All audio clips were traced by the Vocalis Health team of labelers, who also labeled their quality and assured that the interviewee was either positive for COVID-19 (based on the content of the interview) or negative (based on the date of recording). The age and gender of this test set were calculated using the Vocalis Health classifier which was trained previously on 200,000 samples and tested on a hold-out set of 2,800 mutually exclusive samples, reaching an accuracy of 94% for age classification and 99.5% for gender. The baseline characteristics of the free speech test set are summarized in Table 3.

Table 3 - Baseline characteristics of the free speech test set.

	Total participants (n=326)	Positive COVID-19 (N=27)	Negative COVID-19 (n=299)
Age (y)	27±10	31±15	27±10
Males	201 (61.7%)	15 (55.6%)	186 (62.2%)
Language: Hebrew English	 14 (4.3%) 312 (95.7%)	 14 (51.9%) 13 (48.1%)	 0 (0%) 299 (100.0%)

As described above, the first 10 seconds of continuous speech in each recording were used in the feature extraction process, followed by the support vector machine classifier that was optimized on the scripted speech training cohort (Table 2). A threshold of 0.5 was chosen, meaning that each recording with a result above 0.5 was labelled as positive. The biomarker classification results were compared to the COVID-19 labels (see the confusion matrix in Figure 3). Sensitivity of the biomarker was 51.8% [95% CI 32-71%], specificity was 78.3% [95% CI: 73-883%], positive predictive value (PPV) was 17.7% [95% CI: 12.3-24.7%] and negative predictive value (NPV) was 94.7% [95% CI: 92.4-96.4%].

In order to verify that the biomarker is language agnostic, we performed a sub-analysis which included only the English speakers (13 participants) from the positive group (27 participants). The sensitivity and specificity in this sub-analysis were consistent with the sensitivity and specificity in the full test set (53.8% vs 51.8% and 78.3% vs 78.3%, respectively) (figure 3).

Vocal biomarker	COVID19		
	Present	Absent	Total
Positive	True positive **14 (7)**	False positive **65 (65)**	79 (72)
Negative	False negative **13 (6)**	True negative **234 (234)**	247 (240)
Total	27 (13)	299 (299)	

Figure 3. Confusion matrix describing the classification accuracy for the COVID-19 Vocal Biomarker on the free-speech test set. The sensitivity of the biomarker was 51.8% [95% CI 32-71%], specificity 78.3% [95% CI: 73-883%], positive predictive value (PPV) 17.7% [95% CI: 12.3-24.7%] and negative predictive value (NPV) 94.7% [95% CI: 92.4-96.4%]. The numbers in parentheses represent the sub-analysis for English speakers only.

As the threshold can be adapted to different use cases, an additional threshold of 0.35 was defined to suit the cases in which higher sensitivity is required. The biomarker classification results were compared to the COVID-19 labels (see the confusion matrix in Figure 4). The sensitivity of the biomarker was 88.9% [95% CI 70.8-97.7%], specificity was 47.5% [95% CI: 41.7-53.3%], positive predictive value (PPV) was 13.3% [95% CI: 11.4-15.4%] and negative predictive value (NPV) was 97.9% [95% CI: 94.2-99.3%].

Vocal biomarker	COVID19		
	Present	Absent	Total
Positive	True positive **24**	False positive **157**	181
Negative	False negative **3**	True negative **142**	145
Total	27	299	

Figure 4. Confusion matrix describing the prediction accuracy for the Vocalis COVID-19 biomarker compared to actual labels on the free speech test set with a threshold of 0.35. The biomarker classification results were compared to the COVID-19 labels (see the confusion matrix in Figure 4). The sensitivity of the biomarker was 88.9% [95% CI 70.8-97.7%], specificity was 47.5% [95% CI: 41.7-53.3%], positive predictive value (PPV) was 13.3% [95% CI: 11.4-15.4%] and negative predictive value (NPV) was 97.9% [95% CI: 94.2-99.3%].

Analysis 3: Assessing the Vocalis Biomarker Performance vs. Fever Screening

Fever is the most common screening tool used today for COVID-19. Therefore, the performance of the Vocal Biomarker was compared to fever. To create new training and test sets, 11 COVID-19 positive participants whose voice recordings included counting from 1 to 30 were excluded. An additional 11 COVID-19 negative participants matched for age, gender and language, and whose voice recordings included counting from 1 to 30, were also excluded. COVID-19 positive participants from the online survey dataset, whose voice recordings included counting from 50 to 70, were still included in the new training set. The new training set included a total of 134 recordings (a single recording per participant), which contained scripted or free speech. The characteristics of the training set are described in Table 4.

The new test dataset included all voice recordings from the online survey set that were not included in the training set (783 out of 842). Eleven participants in the test set were positive for COVID-19 and 772 were negative. The characteristics of the test set can are described in table 5.

Table 4 - Baseline characteristics of the symptoms training set

	Total participants (n=134)	Positive COVID-19 (N=67)	Negative COVID-19 (n=67)
Age (y)	34±14.4	36±14.7	32.6±14
Males	94 (70.1%)	47 (70.1%)	47 (70.1%)
Language:			
Hebrew	51 (38.0%)	32 (47.8%)	21 (31.3%)
English	81 (60.4%)	35 (52.2%)	46 (68.7%)
Study method:			
Clinical trial	49 (36.6%)	32 (47.8%)	17 (25.4%)
Other	85 (63.4%)	35 (52.2%)	50 (74.6%)

Table 5 - Baseline characteristics of the symptoms test set

	Total participants (n=783)	Positive COVID-19 (N=11)	Negative COVID-19 (n=772)
Age (y)	37+14	35+9.5	37+14
Males	485 (61.9%)	6 (54.5%)	478 (61.9%)
Language:			
Hebrew	323 (41.3%)	6 (54.5%)	317 (41.1%)
English	465 (59.4%)	5 (45.4%)	454 (58.8%)

As described previously, the first 10 seconds of continuous speech in each recording were used in the feature extraction process, followed by the support vector machine classifier that was optimized on the new training cohort (Table 4). A threshold of 0.5 was chosen, meaning that each recording with a result above 0.5 was labelled as positive.

This test set was used to compare the relative ability to detect COVID-19 between the Vocalis biomarker and fever screening. The results of this comparison are described in Figure 5. The sensitivity of the fever screening was 9.1% [95% CI 0.2.4-41%], specificity was 94.6% [95% CI: 92.7-96%], positive predictive value (PPV) was 2.3% [95% CI: 0.36-13.64%] and negative predictive value (NPV) was 98.6% [95% CI: 98.4-99.8%]. The sensitivity of the biomarker was 54.5% [95% CI 23.4-83.2%], specificity was 77.1% [95% CI: 74-78%], positive predictive value (PPV) was 3.3% [95% CI: 1.9-5.6%] and negative predictive value (NPV) was 99.2% [95% CI: 98.4-99.6%].

Vocal biomarker	COVID19		
	Present	Absent	Total
Positive	True positive **6 (1)**	False positive **177 (42)**	183 (43)
Negative	False negative **5 (10)**	True negative **595 (730)**	600 (740)
Total	11 (11)	772 (772)	

Figure 5. Confusion matrix describing the classification accuracy for the Vocalis COVID-19 biomarker compared to actual label on the symptoms test set. The numbers in the parentheses represent fever classification accuracy. The sensitivity of the biomarker was 54.5% [95% CI 23.4-83.2%], specificity was 77.1% [95% CI: 74-78%], positive predictive value (PPV) was 3.3% [95% CI: 1.9-5.6%] and negative predictive value (NPV) was 99.2% [95% CI: 98.4-99.6%

Discussion and Implications

In this study we demonstrated an association between a non-invasive vocal biomarker and the presence of COVID-19. Our data of 1,217 participants (78 positive and 1,139 negative) came from various recording devices (smartphones, PCs and tablets) in diverse natural environments. In order to demonstrate the capability of the vocal biomarker, first we built a balanced dataset (n=156) and achieved an AUC of 69%, concluding that there is a unique vocal biomarker for COVID-19.

Next, we evaluated the ability of the biomarker to run on free-speech recordings. For this we created a new training set of scripted speech recordings and a test set of free speech recordings. This analysis reached a sensitivity of over 50% and specificity of ~80%, thus strengthening the applicability of this biomarker in the general population in natural environments using spontaneous free speech.

The results also demonstrated that the Vocal Biomarker can be modulated to achieve 88% sensitivity in screening for COVID-19. This heightened sensitivity comes with a cost (higher false positive rates) but can provide a very sensitive tool to screen the population and identify infected individuals.

Last, we compared the biomarker to the widely used screening tool of temperature / fever. This analysis showed much better sensitivity for the biomarker (54.5% compared to 9.1%) at a cost of lower specificity (77.1% compared to 94.6%), concluding that vocal biomarker prediction is at least as good as fever screening and outperforms fever in detecting positive individuals in this small sample size.

The results presented herein support the use of the Vocalis Health vocal biomarker as a first-line screening tool in the effort to safely return to normal activities in the face of ongoing infection risk. It provides a non-invasive tool for COVID-19 risk screening in the general population, relying on voice signals which are accessible, cost-effective and do not require invasive tests. Screening COVID-19 based on voice has the potential to accelerate global efforts to recover from the pandemic.

The voice of healthcare

VocalisHealth

Reach us:

www.vocalishealth.com

info@vocalishealth.com

Disclaimer

This whitepaper is provided for informational purposes only. The product is subject to regulatory clearance as applicable.