|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-L-042 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 19-21 May 2021 | |
| **DOCUMENT** | | | | |
| **Source:** | | TU Dresden, Germany | | |
| **Title:** | | New topic group proposal: Using voice as a biomarker in preventing, predicting and monitoring disease | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Dominik Stosik TU Dresden Germany | | Tel: +4915118185511 Email: dominik.stosik@mailbox.tu-dresden.de |

|  |  |
| --- | --- |
| **Abstract:** | An AI-based solution can only be used with full certainty if it has been rigorously evaluated against a set of criteria. The goal of the TG is to use speech as a potential screening tool. Human voice may be used to evaluate health in a significant way. AI voice assistants can convert speech into a vital indicator, allowing for early assessment and prediction of upcoming health problems. Similarly to how temperature indicates fever, voice biomarkers can offer a more complete view of our health. Consider how much precision and coordination of muscles and brain areas are required to generate voice, and how various disorders might impair one's voice and use of language gradually or acutely. There are a number of interesting firms in this field that are making advances in AI speech technology. The really interesting question is to establish algorithms which allow to draw interconnections between speech and other conditions by connecting national databases with a learning algorithm to sort the abundance of data we have access to. |

**1 Overview**

*Please give a general overview of the use case and of the health problem that is being addressed.*

Acoustic alterations occur as a result of changes in health. These acoustic qualities can predict emotional states, moods, stress, anxiety, sadness, dementia, concussions, heart disease, and other illnesses. AI and machine learning can recognize and analyze these speech qualities better than humans due to massive volumes of data. We have a massive amount of data. We also have systems that operate at a very high level (Alexa, Google, …). The problem is that voice as isolated biomarker probably is very good in detecting early onsets of psychiatric disorders, but when it comes to bridging the divide between AI based voice analysis and other health parameters (i.e. temperature), which soon will be recorded in electronic health records, these systems are limited.

The main goal should be to unite several AI prediction systems in a way that slight voice alternations could indicate temperature changes just to give one example.

**2 Relevance**

*Please explain the relevance of the health problem.*

The COVID-19 pandemic has shown that we were not at all prepared for a situation like this. If we only could put more emphasis on the analysis of language and the connection of the parameters calculated from the algorithms with other systems and nationwide databases, there is a possibility that we could predict outbreaks more accurately. Within the scientific literature there is a consensus that this pandemic is just the beginning of a human induced change in the ecosystem, which probably will have influence on the immune system of following generations. It is to be expected that the intervals between epi- and pandemics will shorten and also other pathogens (i.e. bacteria) and the growing bacterial resistance to antibiotics will have a significantly negative impact on the human immune system and will be an additional burden for the healthcare system (more than 2.8 million antibiotic-resistant infections occur in the U.S. each year, and more than 35,000 people die as a result. In addition, 223,900 cases of Clostridioides difficile occurred in 2017 and at least 12,800 people died. Source: CDC’s Antibiotic Resistance Threats in the United States, 2019 (2019 AR Threats Report)).

The reaction time in most cases will be crucial and early detection systems will be needed more than ever. Not only by reacting quickly to a Pandemic, but also by developing methods (AI-based) which can predict outbreaks before they get out of control and lead to a situation that leads to a breakdown of the healthcare system, as we have witnessed at the beginning of the COVID-19 pandemic.

**3 Impact**

*How could an AI solution provide an improvement relative to the current practice (e.g. will it lead*

*to better care, lower cost, or higher efficiency)?*

*Furthermore, describe the potential impact of benchmarking AI-based solutions for the use case.*

By adding another biomarker to already existing AI prediction models there can only be a synergistic effect. By interconnecting voice with other parameters from nation databases, training these models there is a possibility that this addition will lead to a better care, because vaccines could be delivered to a region where we would - based on an AI prediction - expect a higher incidence than in other regions. Logistics and prevention in general lead to better care. Vocal biomarkers, like pulse and blood pressure biomarkers, are medical indications derived from qualities in our voice. Capturing and analyzing small variations in our voice, such as pitch, intensity, vocal tract coordination, jitters, shimmers, tremors, pauses, and so on, results in a variety of health assessments that offer a more complete picture of our health in seconds. When we will be able to react faster than we were in the past, this will decrease costs significantly. We would be able to work more efficient and by being a step ahead vaccines would not be wasted, because of wrong distribution or transport issues.

**4 Existing work**

*Does the project start from scratch?*

*If not, please provide a brief overview of existing work in the area of the project.*

*Specifically, please explain the current state of the art and how the problem is currently addressed.*

Right now start-ups are focussing on monitoring the voice and derive possible worsening of mental diseases. Also these companies are focussing on actively trying to engage the patient to talk to a voice assistant to establish patterns that allow to predict different possibilities of disease progression. In that case this project would not start from scratch. The idea - once more - is to create a framework which works on top of existing AI models, being trained to establish interconnections and uniting all AI based prediction models to one whole system.

**5 Feasibility**

*Is the project feasible, based on the current state of the art? Yes.*

*Please explain in a few sentences how you see the project progressing from start to finish.*

Start: Gather as much data as possible (Voice AI algorithms can be trained by audiobook recordings) or by using the open accessible Mozilla Common Voice Project, sort that data and create a framework which is able to work on top of other models. The idea should be to aim for a border-crossing system by trying to focus on integration and establishing standards in automated workflows while keeping sensitive patient data safe.

**6 Data availability**

*Please describe the data sets that are available for the project.*

Speech Accent Archive: The speech accent archive was established to uniformly exhibit a large set of speech accents from a variety of language backgrounds. As such, the dataset contains 2,140 English speech samples, each from a different speaker reading the same passage. Furthermore, participants come from 177 countries and have 214 different native languages.

Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): RAVDESS contains 24 professional actors (12 female and 12 male), vocalizing the same statements. Not only this, but the speech emotions captured include calm, happy, sad, angry, fearful, surprise, and disgust at two levels of intensity.

TED-LIUM Corpus: The TED-LIUM corpus is made from TED talks and their transcriptions available on the TED website. It consists of 2,351 audio samples, 452 hours of audio. In addition, the dataset contains 2,351 aligned automatic transcripts in STM format.

Google Audioset: This dataset contains expanding ontology of 635 audio event classes and a collection of over 2 million 10-second sound clips from YouTube videos. Moreover, Google used human labelers to add metadata, context and content analysis.

LibriSpeech ASR Corpus: This corpus contains over 1,000 hours of English speech derived from audiobooks. Most of the recordings are based on texts from Project Gutenberg.

*In particular, can you provide data that have not been (and will not be) disclosed?*

We are thinking of establishing partnerships where we get access to (the more and more frequent teleconsultations). It is crucial to have data from real patients, recognize patterns which - I am convinced - will with enough machine learning training at some point be able to predict the temperature based on the results of a meta-algorithm and help the doctor in his work. It aims to be an assistant to the diagnostic process.

These test data will only be used to evaluate the AI solutions.

*Can an example subset of the data be made available to AI developers?*

This is possible in theory. As I am currently working at a university hospital, patients could be informed that consultations via video or phone are recorded and after a process of anonymization stored in a database where AI models could be fed with that data allowing a learning process that will need to be readjusted on a regular basis.

*Please briefly describe the data format and how (if relevant) the data have been labelled/annotated.*

TBC

*Do you see legal obstacles to sharing the data with FG-AI4H (data protection and privacy laws,*

*copyright)?*

No, rather the opposite. By approaching teaching hospitals as a topic group appointed by ITU/WHO the chances that data will be shared are rather high.

*Furthermore, are there open data sets available for training purposes?*

There need to be open data sets for training purposes. Just imagine how a framework, which interlinks different models and adding the component of human voice could change way we could deal with situations like the ongoing pandemic.

*Would you be able to contribute to an open data set for training purposes?*

Yes.

**7 Data quality**

*Please demonstrate that the available data are of high quality.*

We are working on solutions to reduce the noise from real patient / doctor conversations. This works - in a simplified way - by training the model like someone would pronounce something (a lecturer reading an ebook) and comparing that to the real life examples. Changes, which are so subtle, that a human ear would not be able to grasp can be registered by AI.

**8 Annotation/label quality**

*Please demonstrate that the annotations/labels of the data are of high quality.*

TBD

**9 Data provenance**

*Please demonstrate that the data have been obtained in a professional and ethical way.*

*Do the data come from a variety of sources (e.g. clinical environments)?*

The data will come from clinical environments. Hospitals, where doctors are consulting patients at the moment via video- and phone calls. Once the AI algorithm is trained enough we will start to look for interconnections accessing national databases.

**10 Benchmarking**

*Please describe what type of AI model/algorithm you expect to be benchmarked by your test data.*

*How should the AI model/algorithm be evaluated (including statistics/metrics)?*

TBD

**11 Organizer**

The organizer is Dominik Stosik a soon to be Radiology Resident at TU Dresden in Germany, who studied at Wrocław Medical University in Poland. In the next weeks I will have my disputation of my doctoral thesis at TU Dresden at the Institute of Clinical Pharmacology. I am experienced in informatics and interested in Artificial Intelligence in Medicine. Additionally I am member of the ITU & WHO Topic Group: Artificial Intelligence for Health - Radiology.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_