Interactive visualization to validate AI recommendations for COVID-19 diagnostic of CT scans

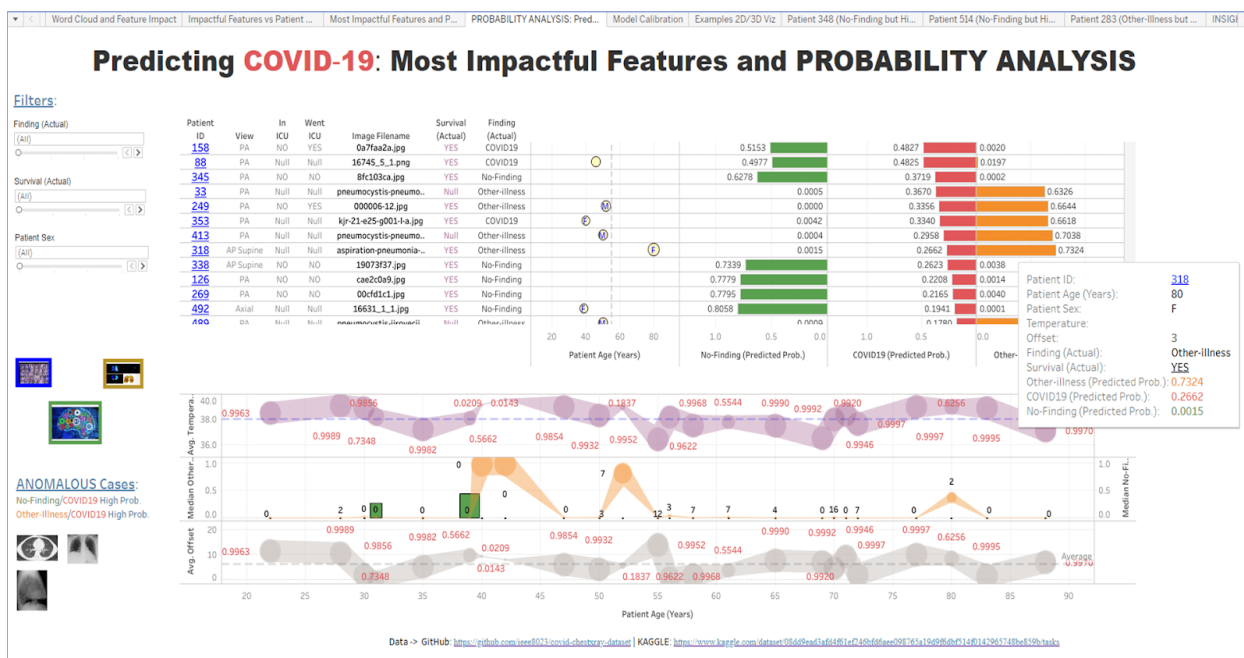Steve Petruzza[1], Ángel Aponte[2], Valerio Pascucci[1]
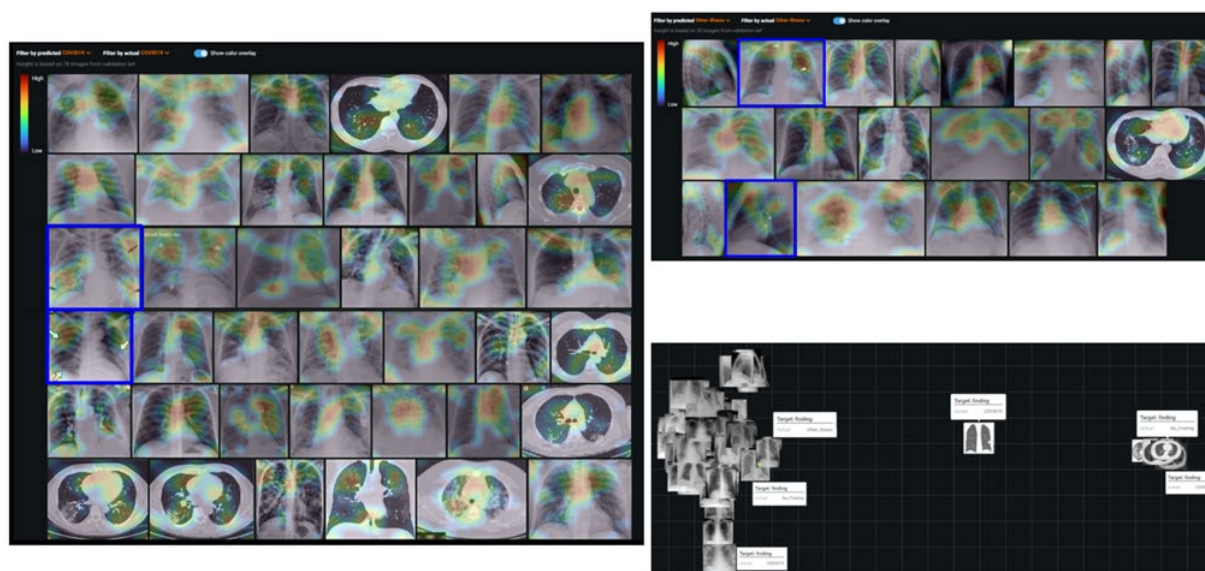
[1]University of Utah, [2]Oilstone

Version 1.0

Distribution: Approved for public release

The goal of this project is to construct an interactive dashboard to visually explore real-time data of suspected COVID-19 patients and inspect actual findings (e.g., whether they were infected or not) and their treatment responses. The platform aims at helping healthcare professionals to identify patterns and anomalies in COVID-19 patients as well as develop new recommendations and best practices.

As reported in an interview [4] in late 2020, "Ángel Alberto Aponte is one such data scientist on the frontlines of COVID-19 research. Through advanced machine learning (ML) techniques and the use of key technologies like Trifacta, Ángel has built a COVID-19 Diagnosis Tool that helps healthcare professionals better respond to COVID-19 cases." The core technology behind the Dashboard is Tableau, since it allows to provide visually detailed information for patients suspected of having COVID-19. With such information, healthcare professionals could identify more quickly cases involving anomalous patterns and deliver proactive recommendations for new cases before they become symptomatic and harder to treat. The following figure shows the main panel of the interactive Dashboard developed to explore treatment response and illness classification.

Interactive visualization to validate AI recommendations for COVID-19 diagnostic of CT scans - CI CoE Pilot, UT, Oilstone
*Distribution: Approved for public release*

1

Integration of Data Visualization and Machine Learning tools is essential to make the tools reliable both in terms of interpretability of the ML models developed and real-time communication of the finding with healthcare professionals. The inference engine behind the Dashboard was built with a machine learning model that derives a classification model with reliable predicting power. Visual inspection of the multi-class grouping of the data (Figure below) is essential to validate the effectiveness of the approach and connect similar cases.

A key component that the Tableau system could not perform interactively was the visual

Interactive visualization to validate AI recommendations for COVID-19 diagnostic of CT scans - CI CoE Pilot, UT, Oilstone
Distribution: Approved for public release

2

exploration of the high resolution images of the individual cases. This is a key functionality for the healthcare providers since it would highlight the pathological regions and allow to better diagnose future cases. The input data come from an heterogeneous collection from the Joseph Paul Cohen Ph.D. project (Postdoctoral Fellow, Mila University of Montreal) that is publicly available on GitHub [2]. It contains a huge variety of data types and structures, such as images of chest X-rays (in jpg or png format) and CT scans, clinical notes (in text format), geographic locations, and other relevant numeric and categorical features. CT scans are stored in Neuroimaging Informatics Technology Initiative (nifti) file format. To overcome the limitations of traditional systems, the data was processed with a python script to transform the original volume into a multiresolution dataset that is more suitable for interactive visualization and data streaming. The web interactive visualization is done using the open source library [3] which includes a WebGL2 ray casting volume renderer that allows to visualize volumes and surfaces in any modern browser without requiring installation of additional softwares. The interactive visualization is made possible by the multiresolution data streaming infrastructure which allows to request data progressively in a coarse to fine resolution hierarchy. This means that the viewer only requests a smaller amount of data (and not the full resolution volume data) to perform the visualization interactively. The viewer also allows users to select different color maps and select a range of values to be rendered. Finally, this tool also allows the visualization of iso surfaces (i.e., surfaces connecting points in the data with the same value) which are useful to explore part of the data that might represent particular tissues or structures. For example, in the case of chest-xrays iso surfaces can help separating the bone structures from the lungs (see interactive visualization [1]).

The following picture depicts the lungs of a patient severely infected with COVID-19. Anomalies are shown in red.
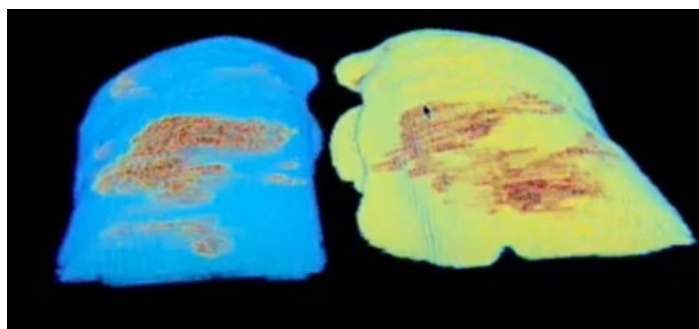
A video of the interactive exploration of those chest x-ray data is also publicly available on YouTube [1]. Furthermore, a tutorial on how to create interactive viewers for volumetric data from binary or image data is available online [5].

The diagnostic tool and Dashboard will continue to be a work in progress as new functionalities will be developed, but some of the components as the interactive visualization tool are now stable and are applicable to a large number of applications in

Interactive visualization to validate AI recommendations for COVID-19 diagnostic of CT scans - CI CoE Pilot, UT, Oilstone
Distribution: Approved for public release

3

histopathology and medical imaging [6].

References

[1] Video interactive visualization of 3D chest x-ray
https://www.youtube.com/watch?v=Tg1p8PsMW2M&ab_channel=OpenViSUS
[2] Joseph Paul Cohen and Paul Morrison and Lan Dao. COVID-19 image data collection, arXiv:2003.11597, 2020. https://github.com/ieee8023/covid-chestxray-dataset
[3] OpenVisusJS: https://github.com/sci-visus/OpenVisusJS
[4] Cariou, Bertrand. 2020. "Predicting COVID-19 Cases with Machine Learning and Trifacta." October 14, 2020. https://www.trifacta.com/blog/predicting-covid19-cases/.
[5] OpenVisus Dataportal tutorial https://wiki.visus.org/index.php/ViSUS_Data_Portal
[6] Multiscale 3-dimensional pathology findings of COVID-19 diseased lung using high-resolution cleared tissue microscopy Guang Li, Sharon E. Fox, Brian Summa, Bihe Hu, Carola Wenk, Aibek Akmatbekov, Jack L. Harbert, Richard S. Vander Heide, J. Quincy Brown bioRxiv 2020.04.11.037473; doi: https://doi.org/10.1101/2020.04.11.037473

Contact for more information

Valerio Pascucci, pascucci.valerio@gmail.com, Steve Petruzza Steve Petruzza steve.petruzza@gmail.com

Version Log

E.G. Version 2.0 updates the timeline to reflect the extension of the engagement through the end of January 2016.

Interactive visualization to validate AI recommendations for COVID-19 diagnostic of CT scans - CI CoE Pilot, UT, Oilstone
Distribution: Approved for public release

4