Reports and Information Visualization

# Data Exploration and Visualization

Radiology reports require demographic information, such as age, sex, ethnicity, education, marital status, and more, as well as clinical information to properly document the patient’s journey throughout the entire cancer screening procedure (Cancer Care Ontario, n.d.). For this reason, the downloaded report will contain the scores for the prediction, the prediction class, as well as all the patient data available within the DICOM file (see sample report below).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Suject ID | sex | age | modality | side | score | pred\_class |
| D1-0820 | F | 65 | MG | R | [0.5 0.5] | Benign |
| D1-0874 | F | 48 | MG | L | [0.49833673 0.50166327] | Malignant |
| D1-0851 | F | 47 | MG | L | [0.4999991 0.5000009] | Malignant |
| D1-0824 | F | 31 | MG | R | [0.5 0.5] | Benign |
| D1-0979 | F | 39 | MG | R | [0.38917866 0.6108213 ] | Malignant |
| D1-0979 | F | 39 | MG | R | [0.50100005 0.49899995] | Benign |
| D1-0991 | F | 28 | MG | R | [0.52085906 0.4791409 ] | Benign |
| D1-0992 | F | 24 | MG | L | [0.5 0.5] | Benign |
| D1-0959 | F | 51 | MG | L | [0.49924758 0.5007524 ] | Malignant |
| D1-0993 | F | 43 | MG | R | [0.49999914 0.5000009 ] | Malignant |
| D1-1002 | F | 39 | MG | R | [0.33589062 0.6641094 ] | Malignant |
| D1-1025 | F | 32 | MG | R | [0.5 0.5] | Benign |
| D1-1012 | F | 45 | MG | R | [0.5 0.5] | Benign |
| D1-1040 | F | 41 | MG | R | [0.49549985 0.5045001 ] | Malignant |
| D1-1126 | F | 49 | MG | R | [0.4649763 0.5350237] | Malignant |
| D1-0978 | F | 40 | MG | R | [0.50048697 0.49951306] | Benign |
| D1-0926 | F | 44 | MG | R | [0.48899266 0.5110073 ] | Malignant |
| D1-0948 | F | 34 | MG | L | [0.5 0.5] | Benign |
| D1-0927 | F | 45 | MG | L | [0.49991018 0.5000898 ] | Malignant |
| D1-0956 | F | 39 | MG | L | [0.5 0.5] | Benign |

**Table 1.** Sample report downloaded from the website

As shown in the above table, the demographics needed to create the radiologist report together with the imaging modality used, the side from which the image was taken as well as the score and the predicted classification are included. No more data has been included due to the HIPAA procedure for anonymization (Office for Civil Rights (OCR, 2012). The key components of the report are the Subject (or patient) ID which will be used by the radiologist or database engineer to associate the prediction with the data found within the hospital’s EHR system so that all doctors can observe the prediction and make conclusions based on whether further testing is required, or if the doctor can proceed towards beginning the surgical procedure or dispatch the patient as the tumor can be considered to be benign. To supplement the downloaded report, the website will provide a dashboard that contains the following format:

Main Dashboard of Tumor Classifier App

Description automatically generated

Figure 1. Main Dashboard of tumor classifier app with charts displaying prediction, sex, laterality, age, and modality.

The current examples above are based on a dataset that takes side-by-side images using the same imaging modality rather than a collection of images using different imaging modalities. As the application becomes more complex, the machine learning model architecture will be used to develop models that are able to appropriately classify the samples with different imaging modalities. Similar case will occur in relation to the sex variable. In this case, the classifier model is only able to classify side-by-side images of breast cancer patients. Because breast cancer is more prominent in women than men, the samples obtained for training the dataset is only female (jenlmat, 2015). Looking forward into the future, the sex category will be used to balance the dataset for training models that diagnose other types of cancer, such as glioblastoma, lung cancer, etc. The opposite can be said in terms of testicular cancer which will only show in men (Kim et al., 2018). Other cancers that are more prominent in males than females are lung, prostate, and colorectal cancer (Kim et al., 2018). In a case study where a radiologist may have to examine hundreds of patients, it may be convenient to observe the types of cancer that the patients are being examined for together with their associated gender to better associate the model’s prediction with the radiologist’s knowledge and experience at diagnosing cancer. In the case that the radiologist observes a higher level of diagnoses of women with prostate cancer when the study contains an even number of men and women (i.e. 50 female patients and 50 male patients), the radiologist can easily see that there may be something amiss with either the model or the study by observing the charts and graphs within the main dashboard. This saves the radiologist time as he would notice this ordeal long after making multiple diagnoses and mitigates the odds that the radiologist may misdiagnose a patient.

The imaging modality used is also another factor that may be able to help one diagnose the patient with cancer. Computed Tomography (CT) scans are the most common and popular among all imaging modalities for diagnosing cancer, but this does not mean that they are the most efficient (Envision Radiology, 2019). Whereas the CT scan is most effective for diagnosing the stage of cancer, Magnetic Resonance Imaging (MRI) techniques help doctors identify whether a tumor is cancerous or non-cancerous (Envision Radiology, 2019). Finally, Ultrasound images allows doctors to precisely locate any unhealthy tissue (Envision Radiology, 2019). Other existing variations of these three imaging modalities, such as mammograms (X-rays targeted at female breasts), or breast MRI, aim to focus solely and specialize in diagnosing very precise types of cancer and are inefficient at diagnosing cancers outside of their target (Envision Radiology, 2019). Hence the imaging modality is necessary for diagnosing cancer based on images, it is easier to diagnose brain tumors using MRI rather than X-rays due to the exposure to radiating energy when creating the images through the use of x-rays as opposed to the non-dangerous magnetic resonance of MRI (Envision Radiology, 2019).

# On-Screen Data Analytics

The purpose of providing the three pie charts is to allow radiologists to observe any out of order predictions based on population size estimates found within the original data set (figure 1). Radiologists analyze approximately one image every three to four seconds with each imaging exam increasing approximately eight times in terms of images to exam from 1999 to 2010 (Staff News Brief, 2015). For this reason, a summary that may provide the radiologist with predictions, together with the sex and age of the patient is required in this dashboard. The imaging modalities are also shown in the form of a pie chart so that the radiologist may be aware of the imaging modalities used to observe the multiple studies. With the information displayed on the dashboard, the radiologist will be able to identify any patterns that may deviate from the standard, such as a high number of malignant cases in adults from the age range 20-40. This would be odd as breast cancer is more common on older women than younger women (Łukasiewicz et al., 2021). Sex is also displayed as breast cancer is more common on women than men (Łukasiewicz et al., 2021). In the case that there may be some studies which might insinuate that a man has breast cancer, the doctor should be able to use the table below the main dashboard to identify the patient and observe other data pertaining to the patient in question. Other reasons for this indication include the number of malignant breast cancer cases (approximately 13%) that will develop throughout a woman’s lifetime (*Breast Cancer Facts and Statistics*, n.d.).

# Reports and Plots

Exporting a report is a very straightforward procedure. Once all the DICOM files have been uploaded to the application. The algorithm will collect the data while making predictions on said data and displaying the results on the main dashboard (figure 2).

Graphical user interface, chart

Description automatically generated

Figure 2. Image of the main dashboard.

Right after the visualizations is the export button which allows the user to export the data table into a csv file.

Table

Description automatically generated

Figure 3. Table containing key patient parameters within the main dashboard.

Although one is not able to download the entire report using a single button, the user is still able to download the independent charts or graphs as pngs. To do this, hover over the graph that the user desires to download and press the camera button at the top left of the chart or graph (this will also be right next to the plotly logo).

Chart, pie chart

Description automatically generated

Figure 4. Pie chart of tumor diagnosis predictions.

Once the image has been downloaded, the user can find the image within the downloads directory.

Graphical user interface, text, application

Description automatically generated

Figure 5. Downloads folder

For other graphs, the user may observe many more buttons on the top left of the graph when hovering over it with the mouse. In this case, the image icon is shown to the left of the zoom icon (figure 6).

Chart, bar chart

Description automatically generated

Figure 6. Bar graph of predictions.

# Sources:

*Breast Cancer Facts and Statistics*. (n.d.). Www.breastcancer.org. Retrieved October 26, 2022, from https://www.breastcancer.org/facts-statistics#section-differences-in-breast-cancer-statistics-by-race-and-ethnicity

Cancer Care Ontario. (n.d.). *Synoptic Radiology Report for Cancer Imaging*. Cancer Care Ontario.

Envision Radiology. (2019, August 7). *Best Scans to Detect Cancer | Envision Radiology*. Envision Radiology. https://www.envrad.com/best-scans-to-detect-cancer/

jenlmat. (2015, March 18). *Breast Cancer Risk Factors*. Rogel Cancer Center | University of Michigan. https://www.rogelcancercenter.org/breast-cancer/detection-and-prevention/breast-cancer-risk-factors#:~:text=Gender%3A

Kim, H.-I., Lim, H., & Moon, A. (2018). Sex Differences in Cancer: Epidemiology, Genetics and Therapy. *Biomolecules & Therapeutics*, *26*(4), 335–342. https://doi.org/10.4062/biomolther.2018.103

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers*, *13*(17), 4287. https://doi.org/10.3390/cancers13174287

Office for Civil Rights (OCR. (2012, September 7). *Methods for De-identification of PHI*. HHS.gov. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

Staff News Brief. (2015, August 12). *The radiologist’s gerbil wheel: interpreting images every 3-4 seconds eight hours a day at Mayo Clinic*. Appliedradiology.com. https://appliedradiology.com/articles/the-radiologist-s-gerbil-wheel-interpreting-images-every-3-4-seconds-eight-hours-a-day-at-mayo-clinic