**Artificial Intelligence-based tools for thyroid ultrasound**

**Project Workflow(プロジェクト計画案)**

**Aim（目的):**

To develop AI tools to assist thyroid analysis, diagnosis and treatment. This involves creating a segmentation map of ultrasound images, and measuring the relevant features of thyroid, and classifying them into different types of cancer.

To measure TI-RADS and review its effectiveness and scope in cancer diagnosis, and further develop a scoring algorithm which can enhance the diagnosis.

**Goals (Extended)（将来像/赴き):**

To carry out a technical study of Ito hospital Data storage system, and the feasibility of installing Dicom base systems for exporting Dicom data directly from hospital.

To make a technical summary of the results and publish them in peer reviewed journals.

**Project Members (会員):**

1. Sparquest Inc. (Data science subsidiary of Azest Inc.)

#Members

1. Itou Hospital

#Members

1. MGH

#Members

**Project Workflow (計画案（暫定的順番)):**

The project is broadly divided into five phases:

1. Workflow setup and Non-temporal Data segmentation & Classification (Time-independent).
2. To integrate the Dicom image data in the existing systems and update the workflow.
3. To evaluate the existing TI-RADS scoring system and improve the scoring model.
4. Temporal Data Segmentation and Classification (Time Dependent).
5. On-site installation of software in existing hardware systems for real-time imaging.
6. **Workflow setup and Non-temporal Data Segmentation & Classification (2019/10-2019/05):**

This phase involves the analysis of time-independent data, and the development of algorithms or future data preparation.

**Workflow Setup:**

* Installation of Nvidia-GPU driver and setting up docker environment for TensorFlow.
* Installation of required python computer vision, visualization and Dicom libraries for image analysis.
* Data curation and Setup of remote/cloud work-station for receiving and securely operating the data from the hospital.

**Non-temporal Data Segmentation & Classification:**

* *Data Collection*:

Focusing on the common Thyroid nodule types, Normal, Benign, Cyst, Papillary Cancer and Solid lesions, about 300 patient data consisting of Ultrasound Images in JPG format, Pathological and clinical data, is collected from the hospital.

Each Patient has a unique serial number, and we receive 2 Ultrasound image/patient in horizontal and vertical direction.

* *Data Annotation:*

Initial samples of data are annotated by the doctors and technicians of Ito hospital.

The large-scale image annotation for segmentation is outsourced to a trusted company in installments.

The output segmentation map from the company is sent for feedback and review to the Ito hospital

* *Data Pre-processing:*

The scale information for the image is calculated and standardized.

The borders and irrelevant portions of the image are cropped.

The image size is standardized across all images for training, without changing the respective aspect ratio.

Data Augmentation is done by mirroring the image along y-axis to double the dataset.

Algorithms are developed for extracting the segmented colored image based on color codes and converted into model output by one-hot vector encoding.

* *Developing Model:*

Algorithms are developed for semantic segmentation based on U-Net architecture.

Semantic Segmentation map is created for the common thyroid nodules.

The length, area and other characteristics of thyroid and nodules are measured.

The model is subsequently trained for classification with additional input data.

* *Repeating the process:*

The process is repeated with the additional input data from Ito hospital for different classes.

1. **To integrate the Dicom image data in the existing systems and update the workflow (2019/12-):**

* This phase involves the investigation of the existing data lifecycle in Ito hospital, to understand the data storage and retrieval mechanism of images.
* New data-warehousing system (DWH) is proposed for accessing the Dicom images and secure handling of medical image data.
* The hitherto developed model is updated to handle Dicom data.
* The metadata, pathological data and clinical data is further used as inputs for enhancing the model.

1. **To evaluate the existing TI-RADS scoring system and improve the scoring model.**

This phase involves data preparation to review the existing TI-RADS system for the collected dataset, and improvement the scoring model using ML algorithms.

* The dataset is prepared based on the current metrics used in TI-RADS.
* The TI-RADS score is calculated along with the relevant statistics, to review the scoring system.
* The dataset is trained using a ML model to calculate the relationship between the metrics and new scoring system is proposed.

1. **Temporal Data Segmentation and Classification (Time Dependent)**
2. **On-site installation of software in existing hardware systems for real-time imaging**