**SNMMI AI Challenge dataset description**

The dataset for the SNMMI AI challenge is provided as an excel sheet including 500+ radiomics features. The sample size of the data contains all features from 296 subjects. The dataset has the following characteristics and information.

The training dataset is derived from the HOVON84 trial in diffuse large B-cell ly,phoma and described in detail in:

Lugtenburg PJ, de Nully Brown P, van der Holt B, D'Amore FA, Koene HR, de Jongh E, et al. Rituximab-CHOP with early rituximab intensification for diffuse large B-cell lymphoma: a randomized phase III trial of the HOVON and the Nordic Lymphoma Group (HOVON-84) J Clin Oncol. 2020 doi: 10.1200/JCO.19.03418.

The actuals dataset of 296 patients is also described in:

 18F-FDG PET Improves Baseline Clinical Predictors of Response in Diffuse Large B-Cell Lymphoma: The HOVON-84 Study. Burggraaff CN, Eertink JJ, Lugtenburg PJ, Hoekstra OS, Arens AIJ, de Keizer B, Heymans MW, van der Holt B, Wiegers SE, Pieplenbosch S, Boellaard R, de Vet HCW, Zijlstra JM; HOVON Imaging Working Group and the HOVON Lymphoma Working Group. J Nucl Med. 2022 Jul;63(7):1001-1007. doi: 10.2967/jnumed.121.262205.

**Dataset description:**

* The first columns is a simple patient ID without any meaning.
* The second column contains the PFS outcome expressed in months. This is the outcome measure to be used as ‘label’ for training, i.e. you are requested to develop a predictive model that can predict outcome best.
* The third columns provided the event tag for the reported PFS. 0 means no event, 1 means death or progressive disease. Thus 0 reflects end of follow-up and patients may have lived longer without disease progression or death
* The 4th columns contains SUVpeak
* The 5th columns contains the total tumor volume in uL
* SUVpeak and volume are provided and disclosed as they represent the standard available metrics in most PET studies and may be used as basic model parameters or else.
* All subsequent columns contain (anonymized) radiomics features, labeled F1, F2, etc
* a single row contains all the data/features of a single patient. There is only one row per patient

Please be aware that the dataset is real life data from an actual patient study and may contain anomalies, such as:

* duplicate features
* constant value features
* features that occasionally contain zero values (missing data).

Part of the challenge is that you need to include strategies in your model development to deal with real life data errors, such as censoring features, combining features etc. You will be asked to fill out a submission form providing as with as much as possible details on how you processed the data prior and during model training (see next remark)

**Some extra remarks**

You are requested to fill out a questionnaire once you are ready to do a final (external and blinded) test of your model, so that we can capture the most important steps that you undertook to develop and train the model. The external testing dataset will be provided after SNMMI has received this fully filled out questionnaire.

**Figure or merit**

Once you have applied your model to the blinded external dataset, your results should be submitted to SNMMI and will be evaluated by the team. Results requested are only the predicted PFS from the blinded external dataset along with the fully filled out data submission and information sheet. The main performance criterion to assess the best performing model will be the **mean squared error between predicted and observed PFS in months to determine the most accurate model with respect to predicting the correct PFS**. **In addition, we will derive the average c-index for 1, 2 and 3 year’s PFS to assess the best classifying model. For the latter case we need you to provide the predicted probabilities for 1,2 and 3 year’s PFS.** Other metrics may be derived by the team for descriptive purposes.

Your data will be used for a publication summarizing the main findings from the challenge. The full data, including reference data associated with the test set cases, is expected to be made publicly available after the publication of the Challenge. Until the official public release of the data, data downloaded from this site or provided to you by the organizers of the challenge may only be used for the purpose of preparing an entry to be submitted for the AI Challenge. The data may not be used for other purposes in scientific studies and may not be used to train or develop other algorithms, including but not limited to algorithms used in commercial products.

After publication of the AI challenge analysis by the organizers, participants are permitted to publish their methods and findings with acknowledgement of the SNMMI AI Taskforce and a reference to the AI challenge publication by the AI challenge organizers. Participants will be informed by the AI challenge organizers when individual participant publication will be permitted.

For more information, awards and price money, please see also the challenge website.