**Task 4.4** **Transforming Medical Images to Medical Reports and Recommendation Mechanism**

(Leader: CERTH) [M10– M34]

Task 4.4 A critical factor of INCISIVE is to perform automatic or semi-automatic annotation of the medical images. This means that the system will be able to extract the report of the diagnosis by “examining” or “looking inside” the image data.

This is done by initially identifying a set of visual descriptors which can represent the visual content as better as it can be. Then, rule-specific data analysis is carried out to fill in the reports under an automatic (or at least with a minimal assistance) way. Predefined report templates are exploited which have been designed by the physicians.

In addition, the contextual conditions of the image data capturing are stored in the database so that terminology on the name of the tumour, its location, and organs affecting to be pre-generated.

The gaps that we need to fill in are the size of the tumour in 3D coordinates, its invasive in neighbouring regions, comparative progress with respect to previous examination, metastatic behaviours and activities as well as the behaviour of the disease to a treatment protocol.

(from Grant Agreement)

**Available data for training (Incisive)**

* Images
* Templates (excel)
* Medical reports?

**How can we retrieve the information to form the medical report?  
tags to predict / fields from templates**

Pre-generated tags:

* name of the tumour
* location
* organs affecting

Gaps to fill in:

* **size.**  of the tumour in 3D coordinates
  + spatial measurement of segmented area on T.1 - T.2 results
  + max tumor diameter (in cm)
    - templates: (breast) (lung) (colorectal)
  + TNM (T: size of tumor) (Tx/Tis/T1/T2/T3/T4)
    - templates: (breast) (colorectal)
  + cT (1/2/3)
    - templates: (prostate)
* **invasive**. in neighbouring regions
  + TNM (N: nodes invasion) (N1/N2/N3)
    - templates: (breast) (colorectal)
  + Stage (0/IA1/IA2/…/IVA/IVB)
    - templates: (lung)
  + cN (0/1)
    - templates: (prostate)
* Pathological lymph-nodes (0/1)
  + templates: (breast)
* **progress.** comparative progress with respect to previous examination
  + spatial comparison with previous examination (if available)
  + grade (1/2/3)
    - templates: (lung) (colorectal)
  + changes in tumor profile (0/1)
    - templates: (breast)
* **metastatic behaviour.**
  + distant metastasis (0/1)
    - templates: (breast) (lung) (colorectal)
  + distant recurrence (0/1)
    - templates: (prostate)
  + TNM (M: metastasis) (0/1)
    - templates: (breast) (colorectal)
* **behaviour to treatment.**
  + response to treatment (CR/PR/PD/SD)
    - templates: (breast) (lung) (prostate)(colorectal)

**Fields Explanation**

TNM staging cancer

Describes the size of the tumor and how far it spreads

T: tumor size Tx/Tis/T1/T2/T3/T4

N: nodes invasion 0/N1/N2/N3

M: metastasis 0/1

Example: 39 yo male colorectal cancer T3N1M1

Stage

0 0 -> TxN0MO

1 IA -> TisN0M0

IB -> T2N0M0

2 IIA -> T1N1M1 / T2N1M0 / T2N0M0

IIB -> T2N1M0 / T3N0M0

3 IIIA -> T1,2N2M0 / T3N1,2M0 / T4N0,1M0

IIIB -> T4N2M0 / T1-4N3M0

4 IV -> anyTanyNM1

Grade

Tumor grade is the description of a tumor based on how abnormal the tumor cells are. It is an indicator of how quickly a tumor is likely to grow and spread

grade 1 – cancer cells that resemble normal cells and aren't growing rapidly

grade 2 – cancer cells that don't look like normal cells and are growing faster than normal cells

grade 3 – cancer cells that look abnormal and may grow or spread more aggressively

Gx/G1/G2/G3/G4

Response to treatment

CR complete response

PR partial response

PD progressive disease

SD stable disease

Pathological lymph-nodes

0/1

A positive result in a lymph node biopsy means there are cancer cells in the sentinel nodes. This means the cancer has started to spread.

**Open-medical Datasets**

* **ChestX-ray14**

From the national institute of health (NIH) clinical centre. It is an open access chest X-ray dataset that includes 112,120 X-ray images with 14 thorax disease labels (atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening, cardiomegaly, nodule, mass, and hernia). These labels were mined from the original radiologist reports. However, the complete text reports are not publicly available.

<https://nihcc.app.box.com/v/ChestXray-NIHCC>

* **CheXpert**

CheXpert is a dataset consisting of 224,316 chest radiographs of 65,240 patients who underwent a radiographic examination from Stanford University Medical Centre between October 2002 and July 2017, in both inpatient and outpatient centers. Included are their associated radiology reports

<https://www.kaggle.com/mimsadiislam/chexpert>

* **Chest X-rays (Indiana University)**

Open-i has a collection of chest X-Ray Images from the Indiana University hospital network. The IU X-ray dataset contains 7,470 images with 3,955 radiology reports. For each report, there could be multiple images. The image has mainly two views Frontal and Lateral view. XML report contains findings, indication, comparisons, and Impressions.

<https://www.kaggle.com/raddar/chest-xrays-indiana-university>