VALORISER SES DONNÉES DE RECHERCHE PAR LA « COOPÉTITION » INTERNATIONALE

Henning Müller, Adrien Depeursinge, Open Science Days 2023, 9.5.2023

- *Scientific challenges: history, platforms, and perspectives*
  Henning Müller (20min + 5min questions)

- *Organizing a data science competition: A checklist illustrated by the HECKTOR challenge*
  Adrien Depeursinge (20min + 5min questions)

- Exercise: organizing a challenge based on your dataset

  - Reflexion by groups of 3 participants (20min)

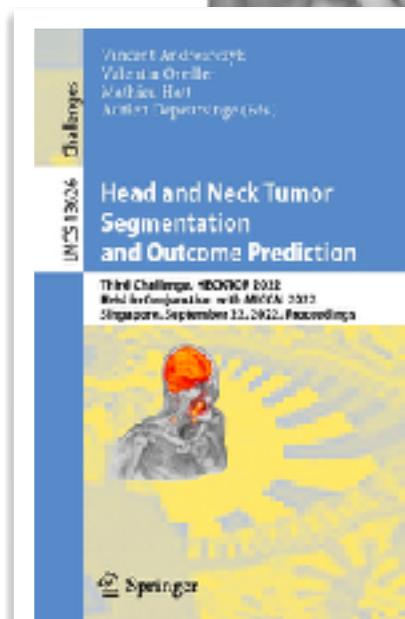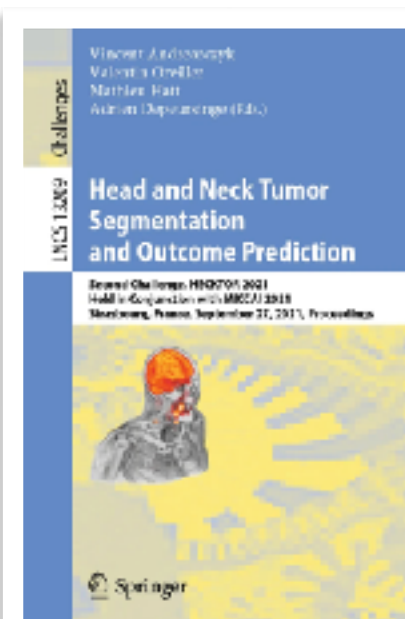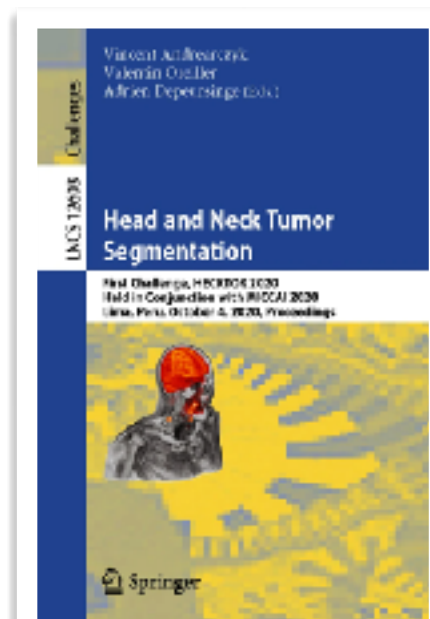 - Short presentations and discussions (20min)

Organizing a data science competition: A checklist illustrated by the HECKTOR challenge

Adrien Depeursinge, Open Science Days 2023, 9.5.2023
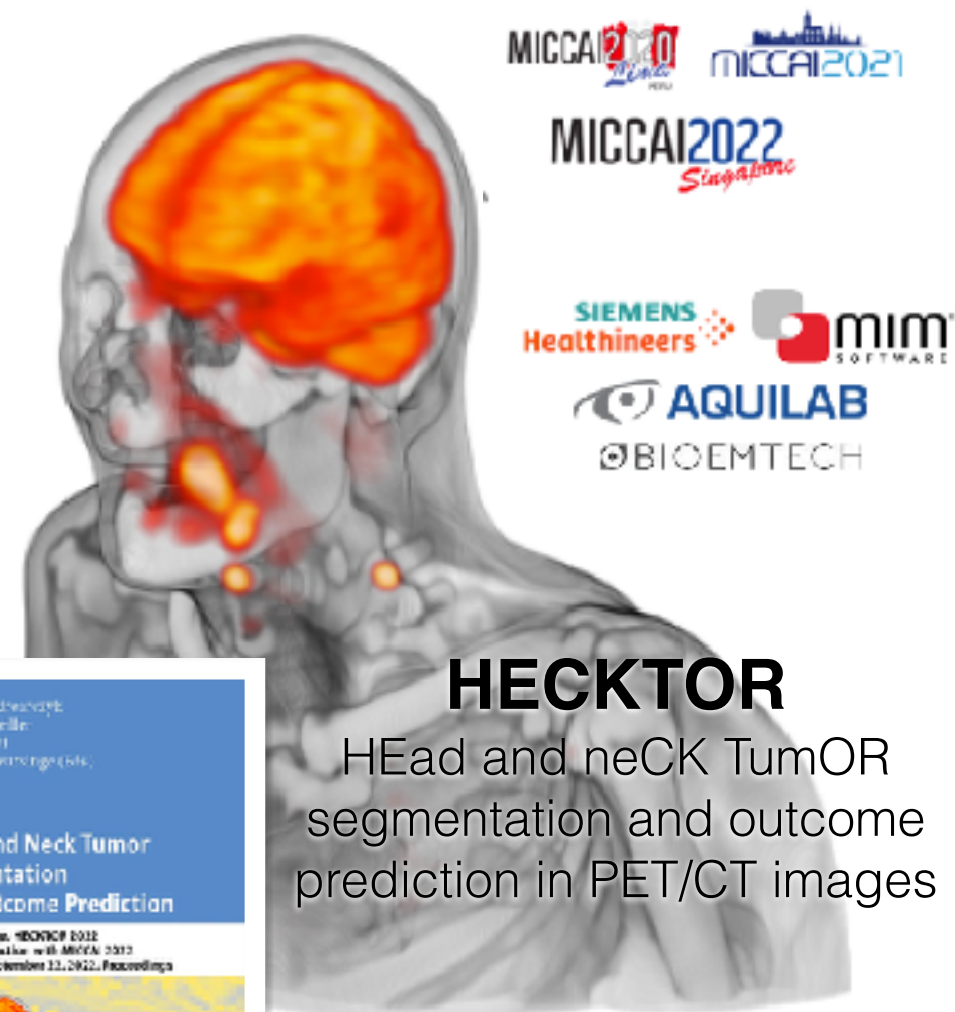
HECKTOR
HEad and neCK TumOR segmentation and outcome prediction in PET/CT images

- Our starting point is a dataset that is either already available or you plan to collect/extend

- We present a generic checklist to assess the relevance and work involved in organizing a data science competition based on a given dataset

  - Exemplify how we addressed the points of the checklist via our experience in leading the organization of the HECKTOR challenge for three consecutive years

- In groups of 3 participants, analyse the potential of organizing a challenge based on an existing/imaginary dataset in terms of the provided checklist

  - 20min analysis

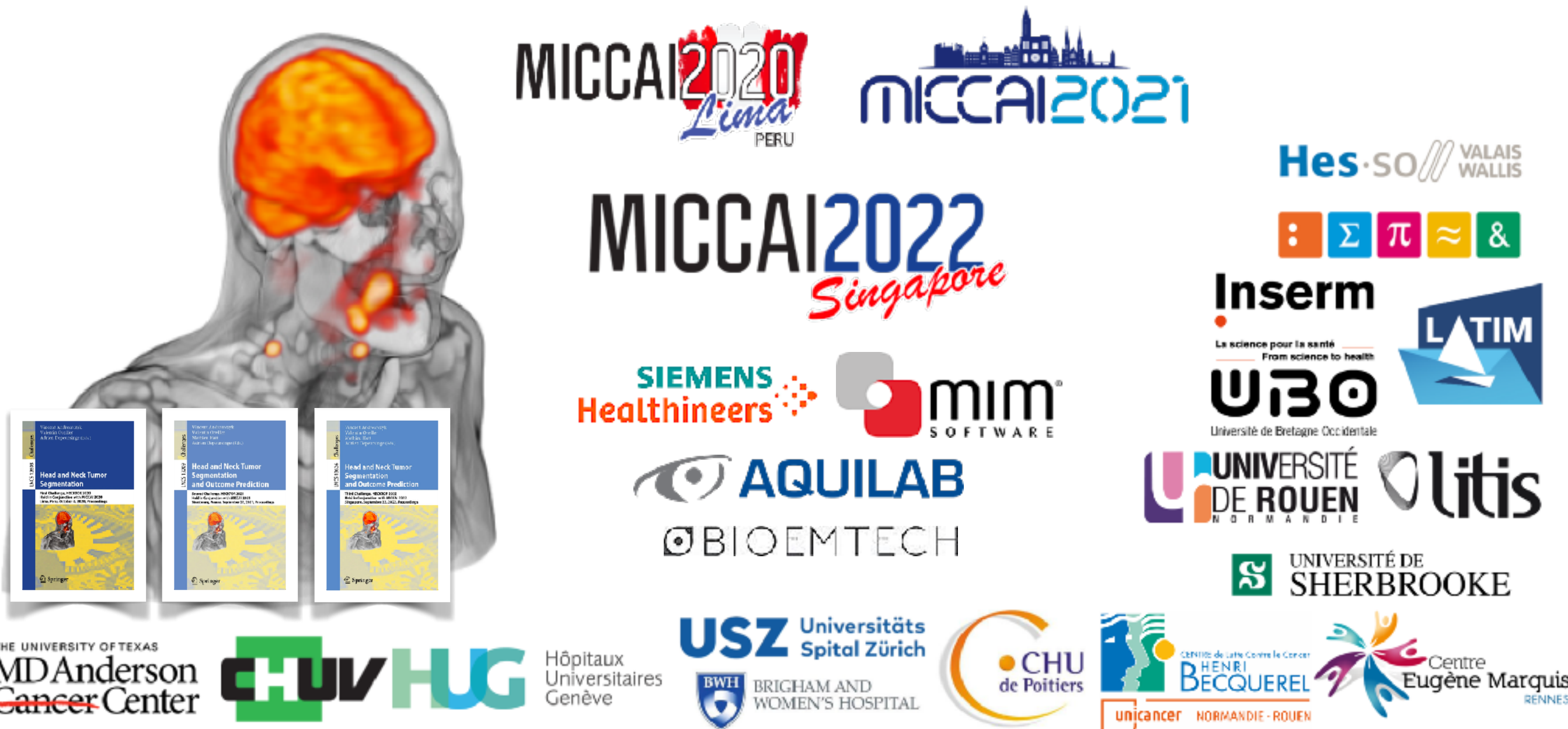 - 20min presentations and discussions

- Based on established guidelines[1] (Maier-Hein et al. 2020)

- Main categories of the checklist

  1. Domain and community

  2. Dataset and curation

  3. Ranking

  4. Challenge organization and implementation

  5. Impact and output

[1] Enhancing the QUAlity and Transparency Of health Research (EQUATOR),
   https://www.equator-network.org/, as of May 2023.
Maier-Hein, L et al. (2020). BIAS: Transparent reporting of biomedical image analysis challenges.
Medical Image Analysis, 66, 101796.

- HECKTOR 2020-2022: HEad and neCK TumOR segmentation and outcome prediction in PET/CT images
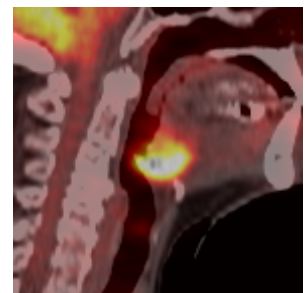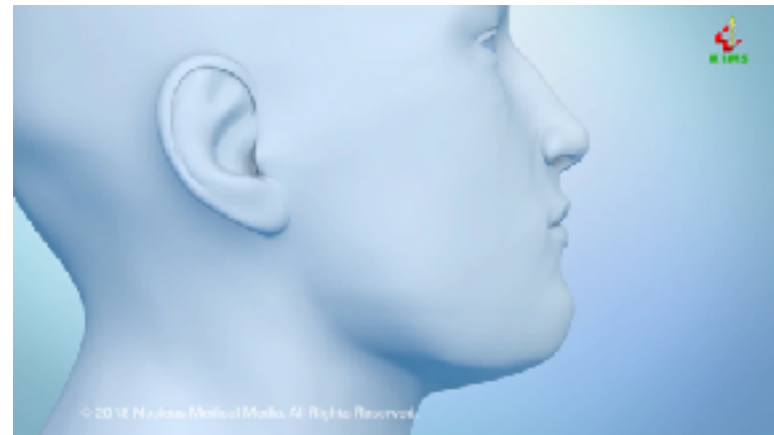
  Oreiller, V et al. (2022) "Head and neck tumor segmentation in PET/CT: The HECKTOR challenge." Medical Image Analysis, 77(1).

- HECKTOR 2020-2022: HEad and neCK TumOR segmentation and outcome prediction in PET/CT images

  Oreiller, V et al. (2022) "Head and neck tumor segmentation in PET/CT: The HECKTOR challenge." Medical Image Analysis, 77(1).
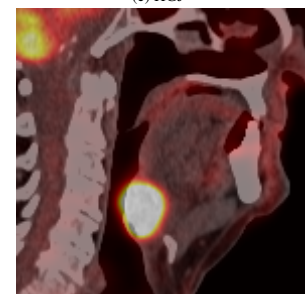
  - H&N cancer 5th leading cancer by incidence (Parkin et al. 2005)

  - High local failure: 40% in first 2 years after treatment (Chajon et al. 2013)

  - Precision oncology: finding optimal treatment for each patient, crucial for patient outcome AND well-being

  - FDG-PET/CT standard for staging and treatment planning

  - **Can AI help predict** the **best treatment** based on PET/CT images and clinical data

    (Vallières et al. 2017, Bogowicz et al. 2017)

    - Correlate visual (lesion size, location and texture) and clinical (age, HPV status, smoking) features with response to treatment

    - **Automatic detection and segmentation** of the primary tumor and met. lymph nodes

    - **Performance is promising but not (yet?) clinically satisfactory**

Parkin DM, et al. (2005) "Global cancer statistics, 2002." CA 55(2).
Chajon E, et al. (2013) "Salivary gland-sparing other than parotid-sparing in definitive head-and-neck intensity-modulated radiotherapy does not seem to jeopardize local control." Rad. Onc. 8(1).
Vallières M, et al. (2017) "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer." Nat. Sci. Rep. 7(1).
Bogowicz M, et al. (2017) "Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma." Acta Oncologica 56(11).

(c) HGJ

(f) CHUP

6

- HECKTOR 2020-2022: HEad and neCK TumOR segmentation and outcome prediction in PET/CT images
  Oreiller, V et al. (2022) "Head and neck tumor segmentation in PET/CT: The HECKTOR challenge." Medical Image Analysis, 77(1).

  - H&N cancer 5th leading cancer by incidence (Parkin et al. 2005)

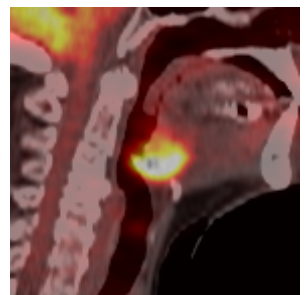 - High local failure: 40% in first 2 years after treatment (Chajon et al. 2013)

  - Precision onc
    treatment for
    patient outco

  - FDG-PET/CT
    treatment pla

Let's organize a challenge to solicit worldwide experts on medical image analysis !
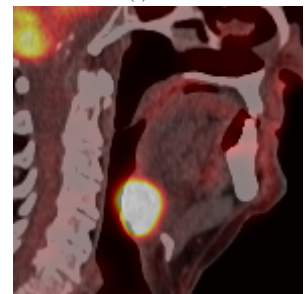
  (c) HGJ

  (f) CHUP

  - **Can AI help predict** the **best treatment**
    based on PET/CT images and clinical data
    (Vallières et al. 2017, Bogowicz et al. 2017)

    - Correlate visual (lesion size, location and texture) and clinical (age, HPV status, smoking) features with response to treatment

    - **Automatic detection and segmentation** of the primary tumor and met. lymph nodes

    - **Performance is promising but not (yet?) clinically satisfactory**

Parkin DM, et al. (2005) "Global cancer statistics, 2002." CA 55(2).
Chajon E, et al. (2013) "Salivary gland-sparing other than parotid-sparing in definitive head-and-neck intensity-modulated radiotherapy does not seem to jeopardize local control." Rad. Onc. 8(1).
Vallières M, et al. (2017) "Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer." Nat. Sci. Rep. 7(1).
Bogowicz M, et al. (2017) "Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma." Acta Oncologica 56(11).

Hes·so// VALAIS WALLIS

7

- ## HECKTOR 2020-2024 challenges comparison

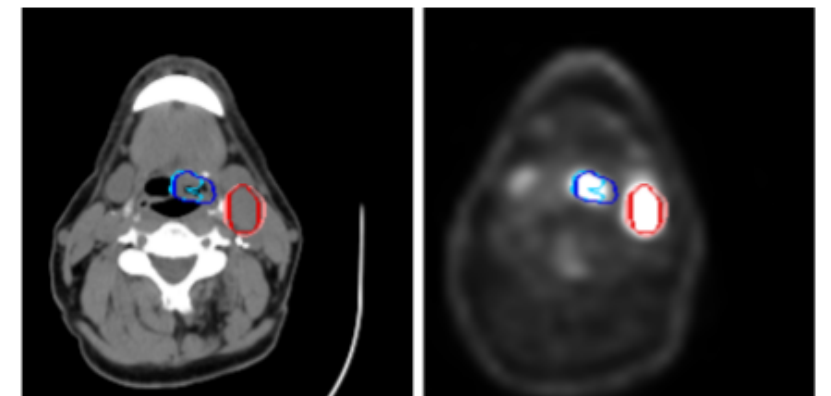|		HECKTOR 2020	HECKTOR 2021	HECKTOR 2022	HECKTOR 2024
Data	# Training subjects	201	224	524	**886**
	# Test subjects	53	101	362	**695**
	# centers	5	6	9	**15**
	Inputs	FDG PET/CT extended oropharyngeal bounding box	FDG PET/CT extended oropharyngeal bounding box	FDG PET/CT full images	FDG PET/CT **RT dose maps** full images
	Clinical data	✅	✅	✅	✅ ++
Tasks	GTVp segmentation	✅	✅	✅	✅
	Outcome prediction		✅ PFS	✅ RFS	✅ RFS
	GTVn segmentation			✅	✅
	HPV status prediction				✅
	Federated learning				✅
	Participant papers	10	31	22	?

- HECKTOR 2020-2022: lessons learned
  - Segmentation of the primary tumor **GTVp** and lymph nodes **GTVn**

Team	DSC_{agg} GTVp	DSC_{agg} GTVn	mean DSC_{agg}	rank
NVAUTO [32]	**0.80066**	0.77539	**0.78802**	1
SJTU426 [41]	0.77960	**0.77604**	0.77782	2
NeuralRad [22]	0.77485	0.76938	0.77212	3
LITO [34]	0.77700	0.76269	0.76984	4
TheDLab [35]	0.77447	0.75865	0.76656	5
MAIA [45]	0.75738	0.77114	0.76426	6
AIRT [46]	0.76689	0.73392	0.75040	8
AIMers [21]	0.73738	0.73431	0.73584	9

- On par with expert performance
- Simple methods work well
  - 3D U-Net always in top three

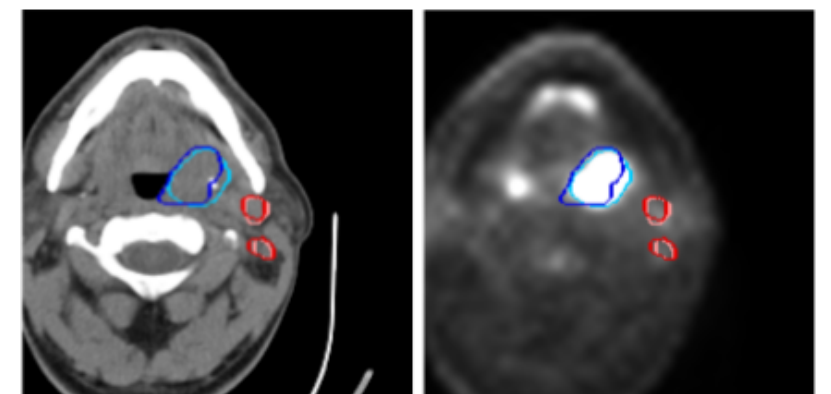|				
RT_UMCG [10]	0.73741	0.65059	0.69400	17
HPCAS [38]	0.69786	0.66730	0.68258	18
ALaGreca [24]	0.72329	0.61341	0.66835	19
Qurit [1]	0.69553	0.57343	0.63448	20
VokCow [30]	0.59424	0.54988	0.57206	21
MLC [43]	0.46587	0.53574	0.50080	22
M&H_lab_NU [40]	0.51342	0.46557	0.48949	23
Average	0.72351	0.68682	0.70517	

Table: Leaderboard segmentation 2022

(a) MDA-203
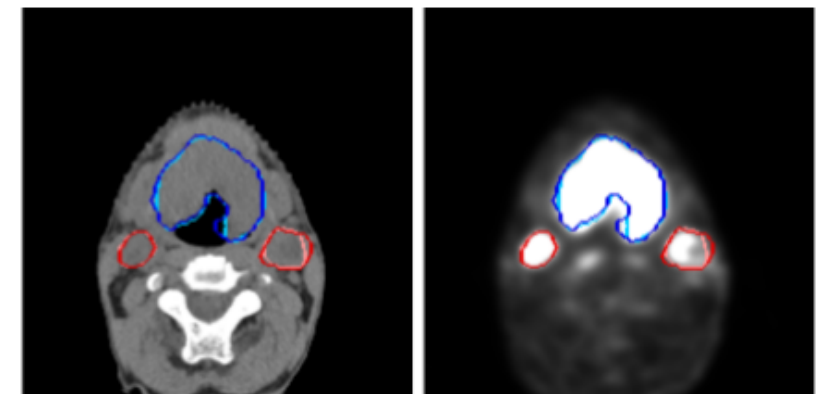
(b) CHB-001

(c) USZ-010

- ## HECKTOR 2020-2022: lessons learned

  - ### Outcome prediction: Recurrence Free Survival (RFS)

Team	C-index	rank
LITO [34]	**0.68152**	1
BDAV_USYD [29]	0.68084	2
AIRT [46]	0.67257	3
RT_UMCG [26]	0.66834	4
RokieLab [49]	0.65817	5
MLC [43]	0.65598	6
VokCow [30]	0.64081	7
junma [25]	0.63896	8
LMU [47]	0.63536	9
TheDLab [35]	0.6305	10
SMIAL [9]	0.61877	11
TECVICO Corp [36]	0.59042	12
Average	0.64769	

Table: Leaderboard
RFS prediction 2022

- RFS prediction not (yet?) ready for clinical use
  - More data needed to better represent (and focus) on subpopulations, e.g. HPV positive only, specific image acquisition protocols, …
- While 4/5 deep learning in top five, the winning team used a very simple radiomics approach

  - Segmentation and outcome prediction tasks are synergistic

    - Learning to segment helps improving outcome prediction (Andrearczyk et al. 2021)
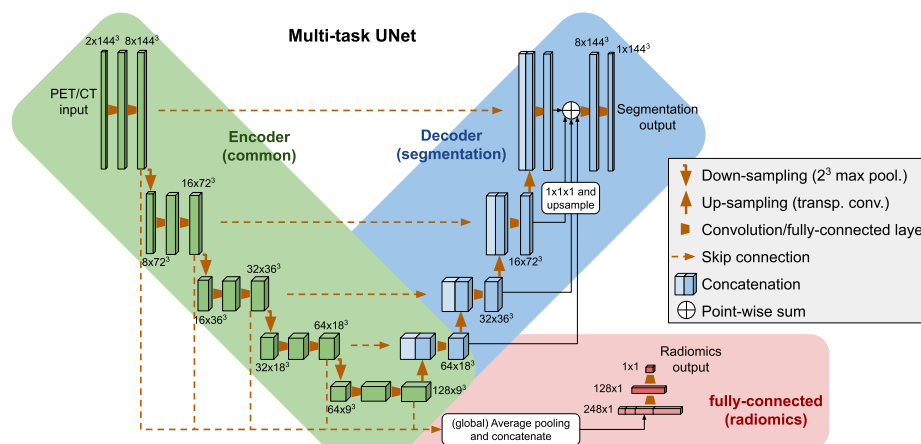
Andrearczyk V et al. (2021) "Multi-Task Deep Segmentation and Radiomics for Automatic Prognosis in Head and Neck Cancer." PRIME.

10

- Based on established guidelines[1] (Maier-Hein et al. 2020)

- Main categories of the checklist

  1. Domain and community

  2. Dataset and curation

  3. Ranking

  4. Challenge organization and implementation

  5. Impact and output

[1] Enhancing the QUAlity and Transparency Of health Research (EQUATOR),
   https://www.equator-network.org/, as of May 2023.
Maier-Hein, L et al. (2020). BIAS: Transparent reporting of biomedical image analysis challenges.
Medical Image Analysis, 66, 101796.

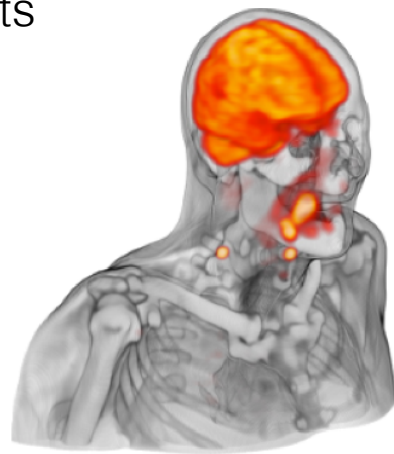1. Domain and community

- Task(s) to solve: motivation and objectives
    - Likely equivalent to the initial goal of the dataset, but not only (reutilization of the data)
        - Scientific (e.g. assess feasibility/maturity, publications as output)
        - Industrial (e.g. assess robustness, commercial product as output)
    - Targeted population
    - State of the art

1. Domain and community

- Task(s) to solve: motivation and objectives
  - Likely equivalent to the initial goal of the dataset, but not only (reutilization of the data)
    - Scientific (e.g. assess feasibility/maturity, publications as output)
    - Industrial (e.g. assess robustness, commercial product as output)
  - Targeted population
  - State of the art
- Disciplines involved: forming a consortium
  - Cover the expertise of all considered aspects for solving the task
    - Task definition, data collection and curation, technology, interpretation of the results

1. ## Domain and community

- Task(s) to solve: motivation and objectives
  - Likely equivalent to the initial goal of the dataset, but not only (reutilization of the data)
    - Scientific (e.g. assess feasibility/maturity, publications as output)
    - Industrial (e.g. assess robustness, commercial product as output)
  - Targeted population
  - State of the art
- Disciplines involved: forming a consortium
  - Cover the expertise of all considered aspects for solving the task
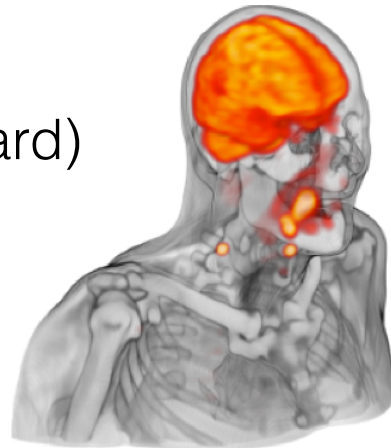   - Task definition, data collection and curation, technology, interpretation of the results

1. Domain and community

- Task(s) to solve: motivation and objectives
  - Likely equivalent to the initial goal of the dataset, but not only (reutilization of the data)
    - Scientific (e.g. assess feasibility/maturity, publications as output)
    - Industrial (e.g. assess robustness, commercial product as output)
  - Targeted population
  - State of the art
- Disciplines involved: forming a consortium
  - Cover the expertise of all considered aspects for solving the task
    - Task definition, data collection and curation, technology, interpretation of the results
- Choice of the venue and format: attract participants
  - International conference
  - Standalone event
  - other (asynchronous leaderboard)

1. Domain and community

- Task(s) to solve: motivation and objectives
  - Likely equivalent to the initial goal of the dataset, but not only (reutilization of the data)
    - Scientific (e.g. assess feasibility/maturity, publications as output)
    - Industrial (e.g. assess robustness, commercial product as output)
  - Targeted population
  - State of the art
- Disciplines involved: forming a consortium
  - Cover the expertise of all considered aspects for solving the task
    - Task definition, data collection and curation, technology, interpretation of the results
- Choice of the venue and format: attract participants
  - International conference
  - Standalone event
  - other (asynchronous leaderboard)

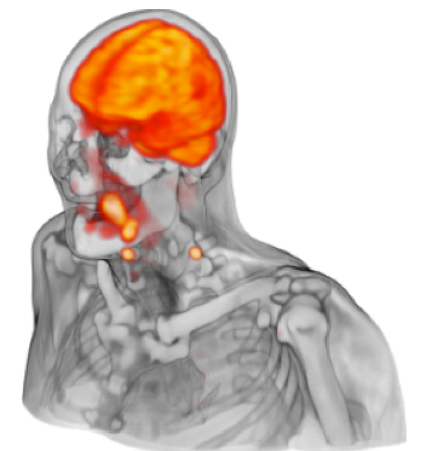## 2. Dataset and curation (Wilkinson et al. 2016)

- **Size** and population
  - Statistical power, risk of overfit (e.g. max 10 features/observation)
  - Sampling the adequate population and representation (diversity)
- **Curation** level
  - Systematic Quality Control (QC), Outliers
  - Unification of the nomenclature (classes, variables, endpoints, metadata)
  - Preprocessing (imposed or free), open source code
  - Versioning and changelog
- Quality of **ground truth**
  - E.g. Single/multiple expert annotator(s), guidelines, crowdsourcing
  - Report potential sources of errors (intra- and inter- annotator variability)

Wilkinson, M. D., et a. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 2016 3:1, 3(1), 1–9.

2. Dataset and curation (Wilkinson et al. 2016)

- Size and population
    - Statistical power, risk of overfit (e.g. max 10 features/observation)
    - Sampling the adequate population and representation (diversity)
- Curation level
    - Systematic Quality Control (QC), Outliers
    - Unification of the nomenclature (classes, variables, endpoints, metadata)
    - Preprocessing (imposed or free), open source code
    - Versioning and changelog
- Quality of ground truth
    - E.g. Single/multiple expert annotator(s), guidelines, crowdsourcing
    - Report potential sources of errors (intra- and inter- annotator variability)

- HECKTOR 2022 : 9 centers, ~900 patients, partially public data
    - Huge effort! (>10 annotators)
    - cloud contouring platform sponsored by mim SOFTWARE
    - Contouring guidelines and outcome definition
      were missing and defined (output)

The Economist
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

2. Dataset and curation (Wilkinson et al. 2016)

- Size and population
  - Statistical power, risk of overfit (e.g. max 10 features/observation)
  - Sampling the adequate population and representation (diversity)
- Curation level
  - Systematic Quality Control (QC), Outliers
  - Unification of the nomenclature (classes, variables, endpoints, metadata)
  - Preprocessing (imposed or free), open source code
  - Versioning and changelog
- Quality of ground truth
  - E.g. Single/multiple expert annotator(s), guidelines, crowdsourcing
  - Report potential sources of errors (intra- and inter- annotator variability)
- Most often, the dataset itself constitutes the main contribution of a challenge!
  - Risk of misleading scientific outcomes

2. Dataset and curation (cont'd)

- Sharing and data protection
  - Anonymization
  - Data Transfer and Use Agreement (DTUA)[1] or licences (e.g. CC-BY[2])
  - Lifecycle
  - Public resources
- Volume and storage (e.g. Zenodo[3], AWS Open Data Registry[4])

[1] E.g. Swiss Personalized Health Network (SPHN) templates in Switzerland, https://sphn.ch/services/dtua/, as of May 2023.
[2] Creative Commons (CC), https://creativecommons.org/about/cclicenses/, as of May 2023.
[3] https://zenodo.org/, as of May 2023.
[4] https://registry.opendata.aws/, as of May 2023.

2. Dataset and curation (cont'd)

- Sharing and data protection
  - Anonymization
  - Data Transfer and Use Agreement (DTU
  - Lifecycle
  - Public resources
- Volume and storage (e.g. Zenodo[3], A

- HECKTOR 2022
  - End User Agreement (EUA) approved by all centers
  - Ethics Committees at center level
  - Switch Drive for sharing (~51GB)

EXHIBITS

Exhibit A: Training and Test collections of Head and Neck Cancer

Institutional Review Boards of all participating PROVIDER institutions permitted use of images and clinical data, either fully anonymized or coded, from all cases for research purposes, only. Retrospective analyses were performed in accordance with the relevant guidelines and regulations as approved by the respective the institutional ethical committees with protocol numbers: MM-JGH-CR15-50 (HGJ, CHUS, HMR, CHUM) and CER-VD 2018-01513 (CHUV). All materials are copyrighted. You must request permission to re-publish any images. At the end of the license agreement the data need to be destroyed.
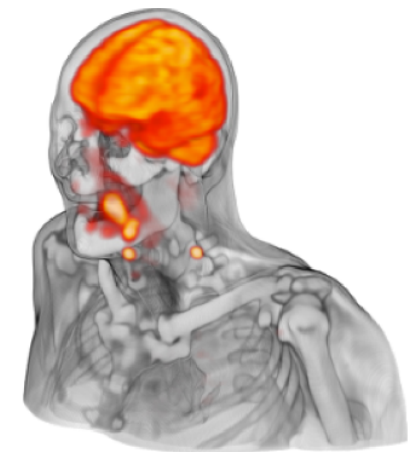
List of PROVIDERS:
HGJ: Hôpital Général Juif, Montréal, CA
CHUS: Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, CA
HMR: Hôpital Maisonneuve-Rosemont, Montréal, CA
CHUM: Centre Hospitalier de l'Université de Montréal, Montréal, CA
CHUV: Centre Hospitalier Universitaire Vaudois, CH

2. Dataset and curation (cont'd)

- Sharing and data protection
  - Anonymization
  - Data Transfer and Use Agreement (DTUA)[1] or licences (e.g. CC-BY[2])
  - Lifecycle
  - Public resources

- Volume and storage (e.g. Zenodo[3], AWS Open Data Registry[4])

- Challenge types
  - Type A: prediction submission where the data "moves" to the algorithm
  - Type B: algorithm submission where the algorithm "moves" to the data

- Validation strategy
  - Train, (validation,) test split(s)
  - Watch out for cheaters! The test set should remain strictly hidden

[1] E.g. Swiss Personalized Health Network (SPHN) templates in Switzerland,
   https://sphn.ch/services/dtua/, as of May 2023.
[2] Creative Commons (CC), https://creativecommons.org/about/cclicenses/, as of May 2023.
[3] https://zenodo.org/, as of May 2023.
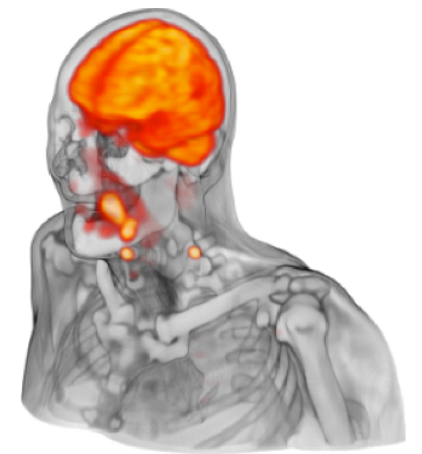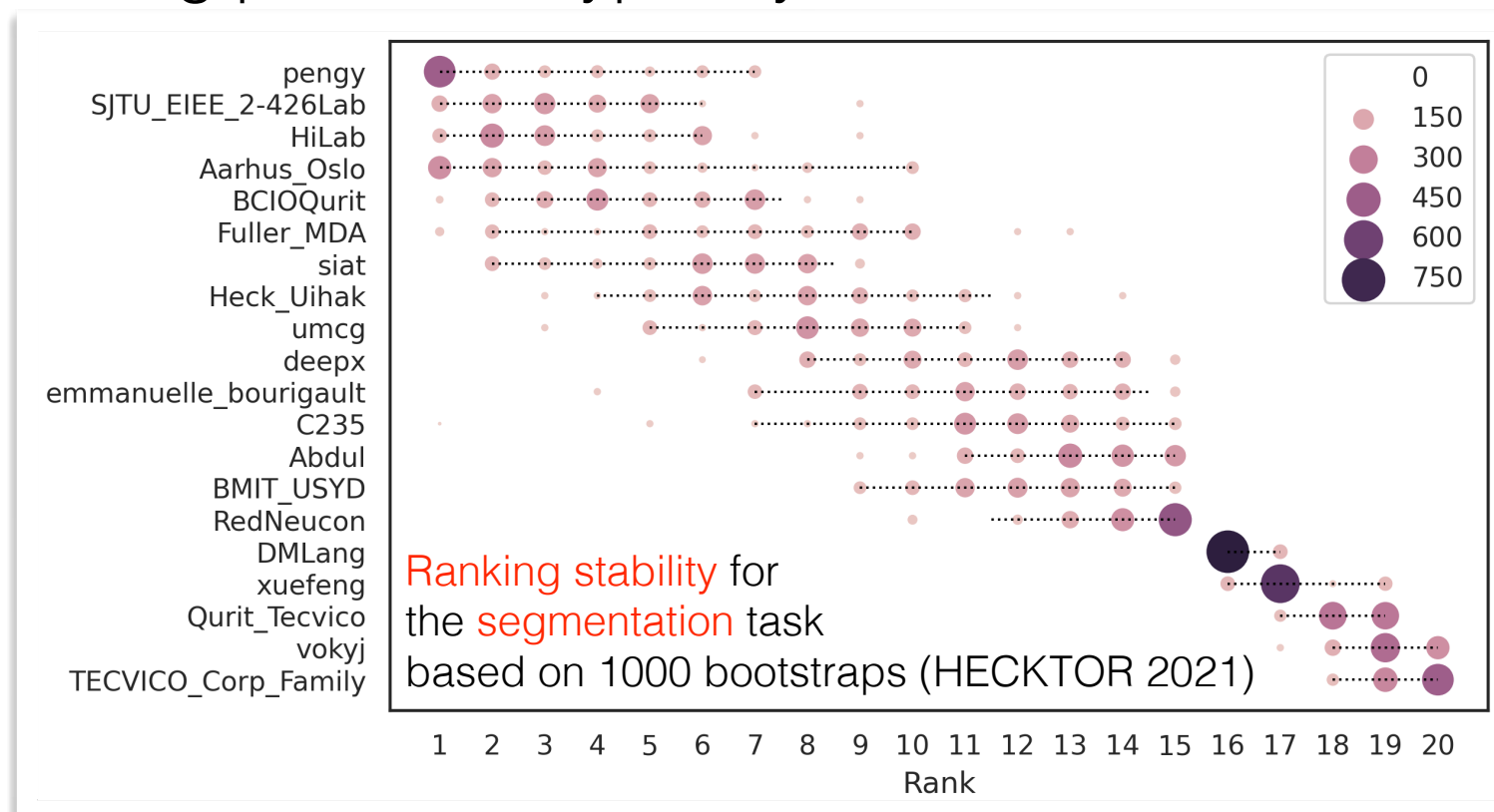[4] https://registry.opendata.aws/, as of May 2023.

3. Ranking

- Performance measure(s) for the considered task(s)
  - Quantitative measures are needed to rank the participants' submissions
  - None are perfect (Maier-Hein et al. 2022)
  - A combination is possible (e.g. average, average rank, Borda-count)
  - Include variability estimation if possible, at a participant/method level
  - Provide evaluation code with example(s) as open source
- Statistical tests and ranking stability (Wiesenfarth et al. 2021)
- One ranking per task is typically considered

Maier-Hein, L. et al. (2022). Metrics reloaded: Pitfalls and recommendations for image analysis validation.
https://arxiv.org/abs/2206.01653v5
Wiesenfarth et al. (2021) "Methods and open-source toolkit for analyzing and visualizing challenge results."
Scientific Reports 11:1 11, 1-15 (2021).

3. Ranking

- Performance measure(s) for the considered task(s)
  - Quantitative measures are needed to rank the participants' submissions
  - None are perfect (Maier-Hein et al. 2022)
  - A combination is possible (e.g. average, average rank, Borda-count)
  - Include variability estimation if possible, at a participant/method level
  - Provide evaluation code with example(s) as open source
- Statistical tests and ranking stability (Wiesenfarth et al. 2021)
- One ranking per task is typically considered

Ranking stability for the segmentation task based on 1000 bootstraps (HECKTOR 2021)

4. Challenge organization and implementation

- Rules
  - Number of submissions allowed
    - Keeping the best or average performance
    - Who/what is considered as submitting team
  - Eligibility criteria for prize and official ranking
    - Must submit a scientific paper describing the approach and/or open source code release
    - Members of the consortium not part of the official ranking
    - Permission to use additional data for training
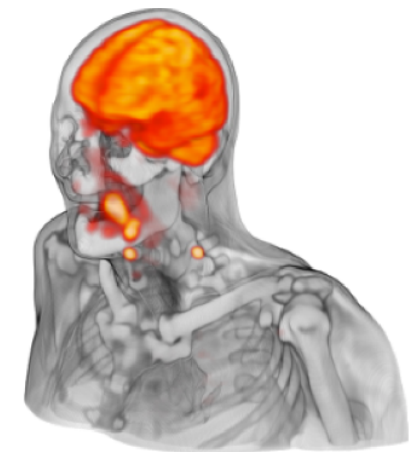   - Semi-automatic or fully automatic only
  - Intellectual property
- Timeline and phases
  - Training, test and algorithms/predictions submissions
  - Scientific session
  - Paper submission deadline and review phase

4. Challenge organization and implementation

- Rules
  - Number of submissions all
    - Keeping the best or averag
    - Who/what is considered as
  - Eligibility criteria for prize a
    - Must submit a scientific pap
      and/or open source code re
    - Members of the consortium
    - Permission to use additiona
    - Semi-automatic or fully automatic only
  - Intellectual property
- Timeline and phases
  - Training, test and algorithms/predictions submissions
  - Scientific session
  - Paper submission deadline and review phase

Timeline HECKTOR 2022:

- May 24th, 2022: **Registration** to the challenge opens
- ~~June 1st, 2022~~ June 7th, 2022: Release of the **training cases**.
- August 1st, 2022: Release of the **testing cases**.
- August 26th to ~~September 2nd~~ September 5th (23:59 PT), 2022: Challenge submissions.
- ~~September 2nd~~ September 5th (23:59 PT), 2022: Paper abstract submission deadline easychair. The abstract can be modified later for final submission).
- September 8th (23:59 PT), 2022: Full paper submission deadline.
- September 8th to October 28th, 2022: Paper review phase.
- September 22, 2022: MICCAI event and release of challenge ranking.

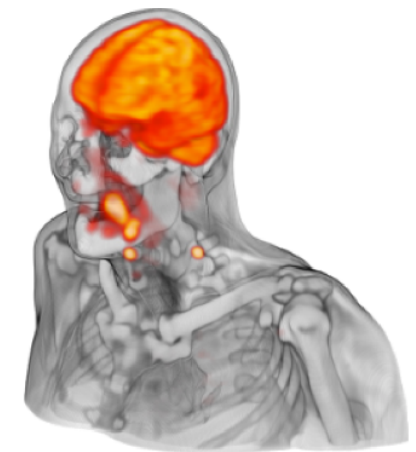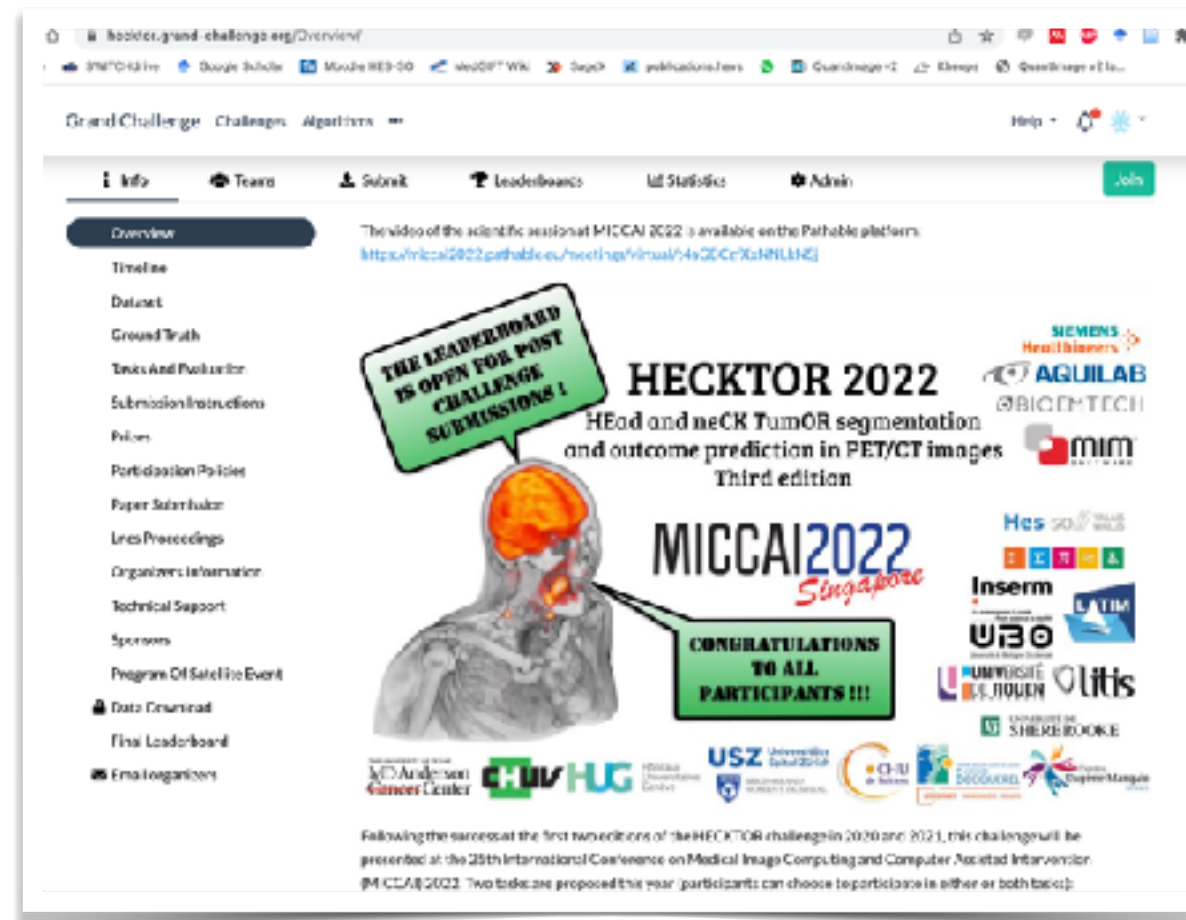4. Challenge organization and implementation (cont'd)

- Prize
  - Money via sponsoring
  - Authorship on the overview paper
- Choice of the hosting platform (e.g. Kaggle, AIcrowd)
  - Often not free, computing fees (evaluation code and/or algorithms for Type B)
  - Implementation of the automated evaluation and ranking
  - Creating and maintaining web pages

4. Challenge organization and implementation (cont'd)

- **Prize**
  - Money via sponsoring
  - Authorship on the overview paper
- Choice of the hosting platform (e.g. Kaggle, AIcrowd)
  - Often not free, computing fees (evaluation code and/or algorithms for Type B)
  - Implementation of the automated evaluation and ranking
  - Creating and maintaining web pages

HECKTOR 2022
Platform/website:
grand-challenge.org

4. Challenge organization and implementation (cont'd)

- **Prize**
  - Money via sponsoring
  - Authorship on the overview paper
- Choice of the hosting platform (e.g. Kaggle, AIcrowd)
  - Often not free, computing fees (evaluation code and/or algorithms for Type B)
  - Implementation of the automated evaluation and ranking
  - Creating and maintaining web pages
- Scientific session: attract attendants
  - Introductory talk(s) by organizers
  - Keynote speaker(s)
  - Oral/poster presentations (e.g. top ranked, most innovative)
  - Award ceremony

4. Challenge organization and imp

- **Prize**
  - Money via sponsoring
  - Authorship on the overview paper
- Choice of the hosting platform (e.g. Ka
  - Often not free, computing fees (evaluation
  - Implementation of the automated evaluatio
  - Creating and maintaining web pages
- **Scientific session**: attract attendants
  - Introductory talk(s) by organizers
  - Keynote speaker(s)
  - Oral/poster presentations (e.g. top ranked, most innovative)
  - Award ceremony

Introductory talk by organizers
11:50 - 12:20: The HECKTOR 2022 challenge, *Vincent Andrearczyk, Valentin Oreiller, Martin Vallières, Clifton Dave Su Ruan, Pierre Decazes, Pierre Vera, Leo Capriolli, Habib Zaidi, Stephanie Tanadini-Lang, Agustina La Grace, Pan Ohara Tankyevych, Dimitris Visvikis, Hesham Elrakawani, Joël Castelli, Ricardo Dias-Correia, Sarah Boughdad, Mari*

Oral Session 1: Automatic segmentation of primary tumors and lymph nodes (Task 1), chair: Kareem Wahid
12:20 - 12:35: A General Web-based Platform for Automatic Delineation of Head and Neck Gross Tumor Volumes in Weiguo Lu
12:35 - 12:50: A Coarse-to-Fine Aggregation Framework for Head and Neck Tumor and Lymph Segmentation in CT a
12:50 - 13:05: Automated head and neck tumor segmentation from 3D PET/CT, *Andriy Myronenko, Md Mahfuzur R.*

Break
13:05 - 13:20

Keynote
13:20 - 14:05: *Radiomics: History and To-Do List, Philippe Lambin*

Oral Session 2: RFS outcome prediction (Task 2), chair: Stephanie Tanadini-Lang
14:05 - 14:20: Deep learning and radiomics based PET/CT feature extraction combined with automatic tumor segme patients, *Baoqiang Ma, Yan Li, Hung Chu, Wei Tang, Luis Luis Ricardo De la O Arévalo, Jiapan Guo, Peter van Coff Dijk and Nanna Maria Sijtsema*
14:20 - 14:35: Radiomics-enhanced Deep Multi-task Learning for Outcome Prediction in Head and Neck Cancer, *Min*
14:35 - 14:50: Head and Neck Tumor and Lymph Node Segmentation and Outcome Prediction from 18F-FDG PET/C, Escobar, Fahad Khalid, Kibrom Girum and Irène Buvat
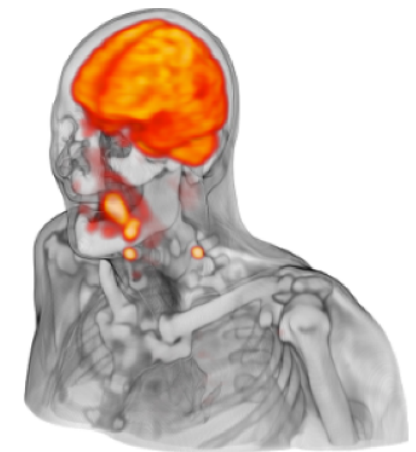14:50 - 15:05: Recurrence-free Survival Prediction under the Guidance of Automatic Gross Tumor Volume Segmenta *Michael Dohopolski, Tao Peng, Weiguo Lu, You Zhang and Jing Wang*

Winners and Awards
15:05 - 15:15: *Valentin Oreiller*

Closing remarks: Feedback from participants / What next
15:15 - 15:20: *Mathieu Hait*

4. Challenge organization and implementation (cont'd)

- Post-challenge activities
  - Post-analyses papers (e.g. inter-algorithm stability, subpopulation analysis)
  - Open leaderboard
  - Super ensembles
  - Open sharing of algorithms and/or predictions (outputs)
  - Preparing the future edition(s)
    - Notoriety: participants are likely to come back with new ideas
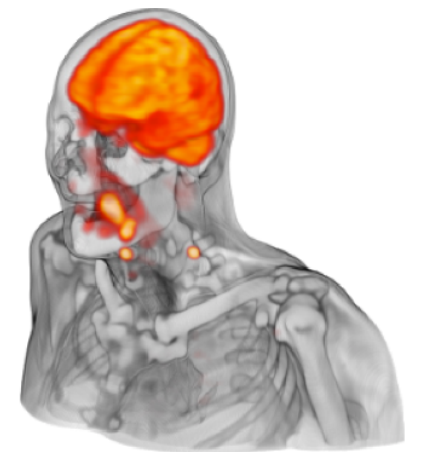- Workforce
  - For all of the above, according to the timeline
  - Team of annotators, one person responsible per task, team of reviewers
  - Governance of the consortium, management

4. Challenge organization and implementation (cont'd)

- Post-challenge activities
  - Post-analyses papers (e.g. inter-algorithm stability, subpopulation analysis)
  - Open leaderboard
  - Super ensembles
  - Open sharing of algorithms and/or predictions (outputs)
  - Preparing the future edition(s)
    - Notoriety: participants are likely to come back with new ideas

- Workforce
  - For all
  - Team of reviewers
  - Gover

5. Impact and output

- For the community
  - Assess the feasibility and maturity of the task
  - Identify problems and potential solutions
    - Ideas for the next edition
    - Motivates the collection of additional data representing specific subpopulations
  - Define standards and guidelines
    - New metrics
    - New guidelines for establishing ground truth
  - Publication of the dataset itself as a resource E.g. (Gatidis et al. 2022)
  - Importance to promote scientific excellence and innovation instead of pure quantitative performance

- For the organizers
  - Strong bibliometric impact (citations)
  - Network and visibility (future projects and funding)
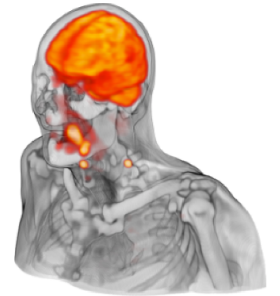 - Fantastic adventure and unique learning process

Gatidis, S. et al. (2022). A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. Scientific Data 2022 9:1, 9(1), 1–7.

5. Impact and output

- For the community
  -
  -
  -
  -
  - Importance to promote scientific excellence and inno
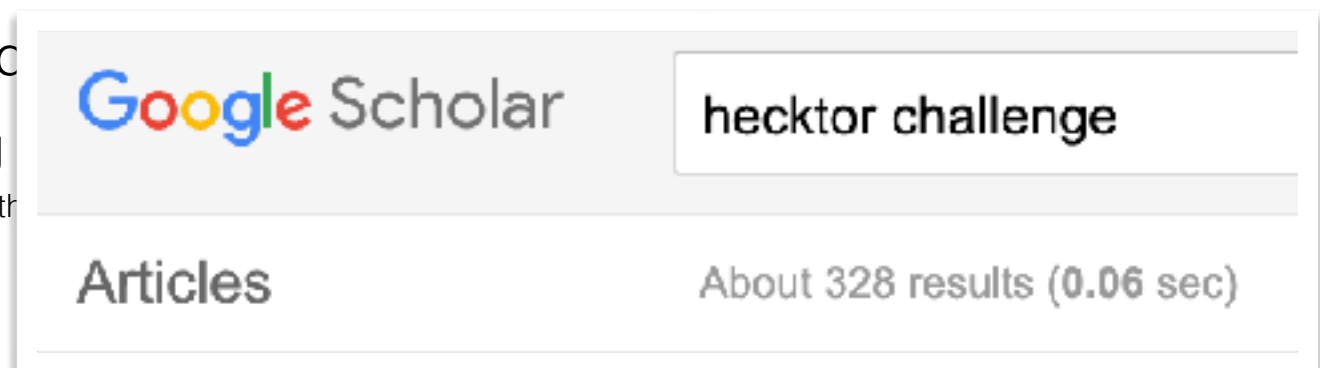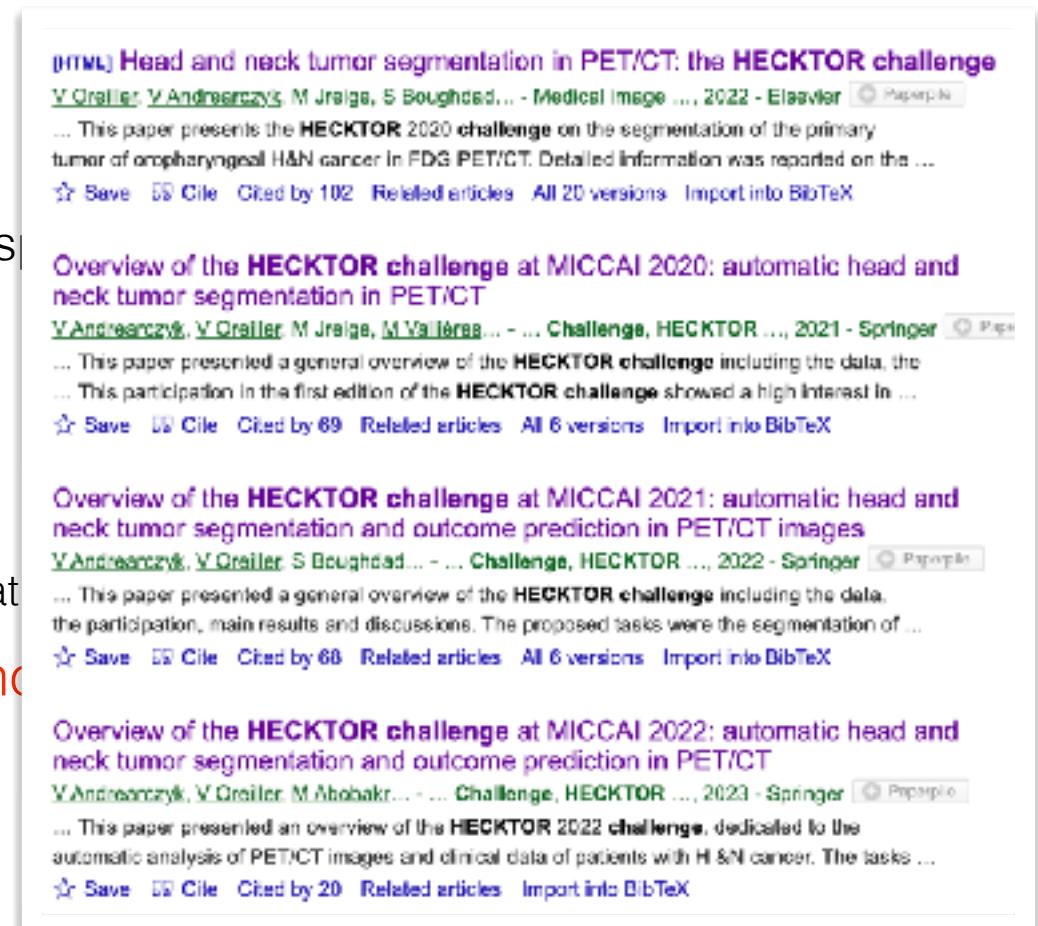   instead of pure quantitative performance

- For the organizers
  - Strong bibliometric impact (citations)
    HECKTOR: 3 proceeding volumes, ~10 papers from us, ~300 citations as of May 2023
  - Network a   lit   p   nd
  - Fantastic   e   u   g

- Conclusions
  - A challenge can have a major impact on the scientific/technological maturity of both the field and the researcher
    - Unique and exciting experience !
  - The dataset often constitutes the most important outcome or resource of the challenge
    - A challenge can highly valorize an existing dataset
    - In the context of the HECKTOR challenge, data curation was among the top time consuming task
    - Starting out from an existing and highly curated dataset can be extremely valuable
  - Organizing a challenge requires taking several decisions that will have a direct impact on the field
    - Poor decisions and data quality can lead to erroneous scientific conclusions
  - Organizing the challenge in the right conference and for 2-3 consecutive years will likely lead to optimal impact

- Simple checklist for assessing efforts related to the organization of a challenge based on your dataset (idea)

  1. Domain and community
     Task(s) to solve, disciplines involved, choice of the venue

  2. Dataset and curation
     Size and population, curation level, availability/quality of ground truth, sharing, volume and storage, validation strategy

  3. Ranking
     Performance measure(s) for the considered task(s)

  4. Challenge organization and implementation
     Rules, timeline and phases, prize, scientific session, post-challenge activities, workforce

  5. Impact and output
     For the community, for the organizers