**4.7 ML meets biomedical Informatics**

Prof William Hsu (UCLA) @ucalwillhsu

[whsu@mednet.ucla.edu](mailto:whsu@mednet.ucla.edu)

<https://hsu-lab.com>

**Learning objectives**

* State the relationship between machine learning and the field of biomedical informatics
* Describe integration approaches for turning multimodal data into knowledge
* Identify barriers in transforming knowledge into practice

**What is biomedical informatics?**

* Biomedical informatics (BMI) is the interdisciplinary field that studies and pursues the effective use of biomedical data, information, and knowledge for scientific inquiry, problem solving, and decision making, motivated by efforts to improve human health.
  + IMIA <https://imia-medinfo.org/wp/>
  + AMIA <https://amia.org/>
* Goal: augmenting human knowledge
  + Not to replace humans

**Subfield of biomedical informatics**

| Basic research |  | | Applied research informatics |
| --- | --- | --- | --- |
| Biomedical informatics:  Methods, technologies, theories |  | | Bioinformatics  *Molecular and Cellular Processes* |
| Imaging informatics  *Tissues and Organs* |
| Clinical informatics  *Individuals (Patients)* |
| Public health informatics  *Populations and Society* |

**Data <-> Practice cycle**



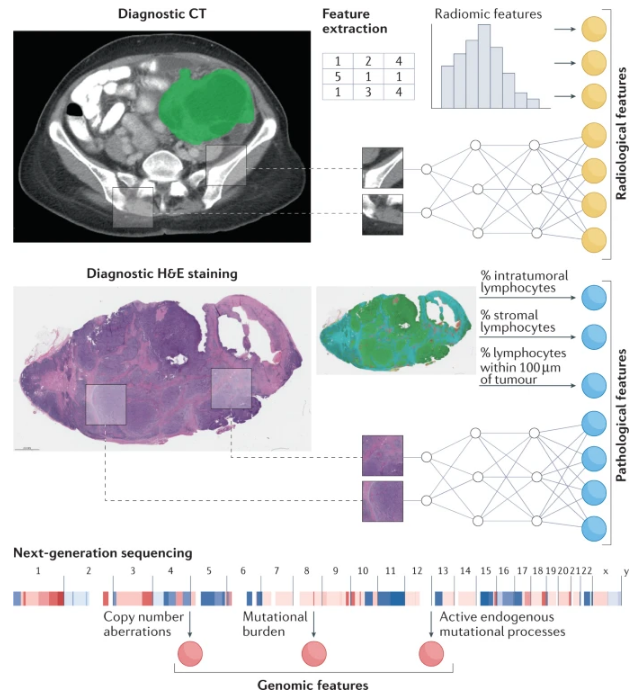
* Chen ES <https://link.springer.com/chapter/10.1007/978-3-030-70558-9_2>

**Towards precision oncology**

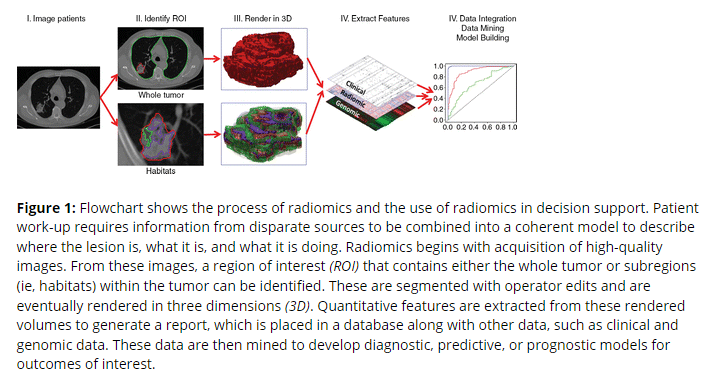
* Diagnosis process is increasingly multi-modal:

1. Patient Presents with Symptoms
2. PCP Orders Imaging Exam
3. Imaging Study Performed
4. Radiologist Interpretation Rendered
5. Suspicious Mass Notes, Biopsy Ordered
6. Biopsy Taken, Analyzed by Pathology
7. Pathology Result Rendered
8. Downstream Care Provided

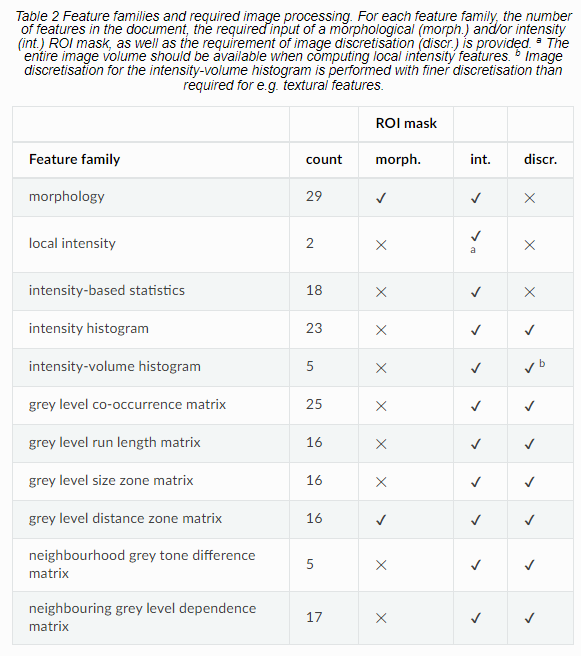
* **Extracting features from multimodal data**



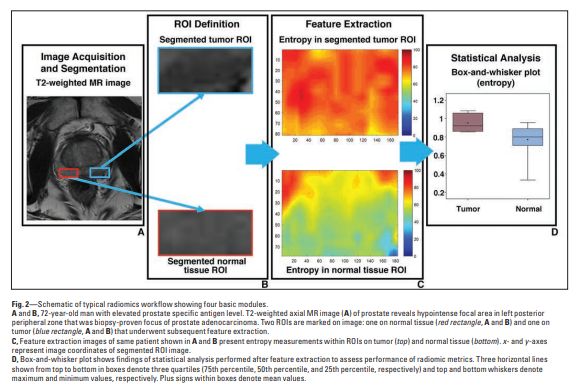
* + Boehm KM et al.
    - [https://dx.doi.org/10.1038/s41568-021-00408-3](https://dx.doi.orq/10.1038/s41568-021-00408-3)
* Radiomics features



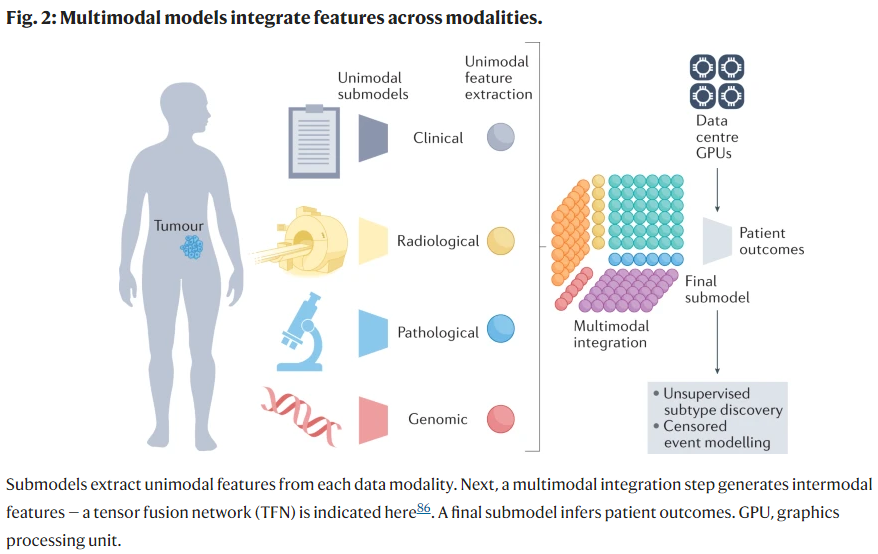
* + Gilies RJ et al., 2015
    - <https://pubs.rsna.org/doi/full/10.1148/radiol.2015151169>
  + Families of radiomic features:



* + More info: <https://ibsi.readthedocs.io/en/latest/03_Image_features.html>
  + Required input of a morphological and/ or intensity ROI mask
  + Requirement of image discretization (discr.)
* Textures:



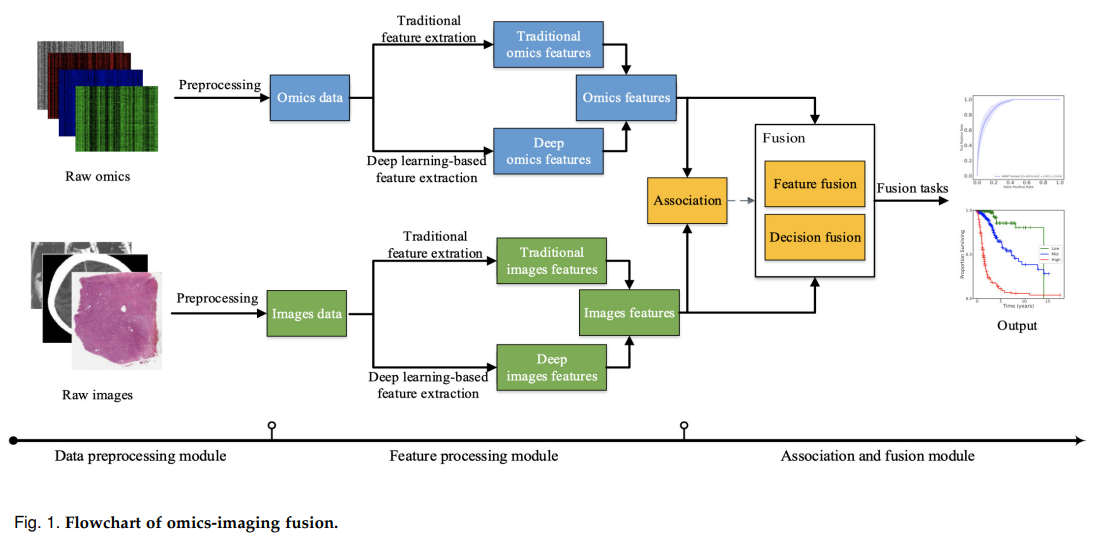
* + <https://www.ajronline.org/doi/pdf/10.2214/AJR.18.20624>
* Extracting features from multimodal data: biomedical data are multi-scale, multi-modal



* + Boehm KM et al.
    - [https://dx.doi.org/10.1038/s41568-021-00408-3](https://dx.doi.orq/10.1038/s41568-021-00408-3)

**Basic types of multimodal integration**

* Feature fusion:
  + Combine features across sources to generate a single feature vector for classification
* Decision fusion:
  + Classify features from each source independently then aggregate the results
* More information: <https://ieeexplore.ieee.org/document/5192987>
* An example of general approach:
  + <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9684949>



| Fusion level | Fusion type | Fusion method |
| --- | --- | --- |
| Feature fusion | Fusion through association | Canonical correlation analysis |
| Independent component analysis |
| Nonnegative matrix factorization |
| Direct fusion | Vector concatenation |
| Sparse representation |
| Multiple kernel learning |
| Deep neural network |
| Graph-based network |
| Decision fusion | Ensemble learning | Random forest |

* Example: Prostate cancer (PCa) aggressiveness
  + Identifying patients who are likely to experience biochemical recurrence (BCR) after radical prostatectomy by fusing radiomic features extracted from the lesions shown in T2w and ADC
  + Existing works:

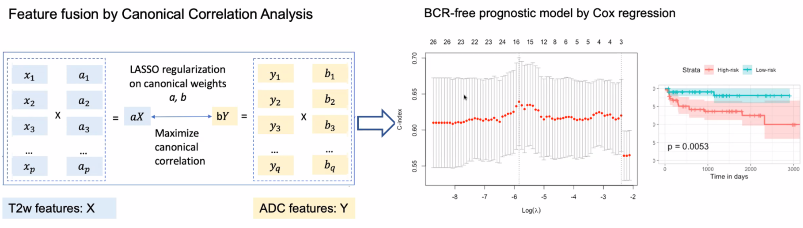
| Name | Data | Treatment | Imaging | Radiomics | Feature fusion |
| --- | --- | --- | --- | --- | --- |
| Gnep et al., 2017  <https://onlinelibrary.wiley.com/doi/10.1002/jmri.25335> | N=47  1 center | RTx | T2w, ADC | 144 intensity,  Texture features | Direct concatenation |
| Shiradkar et al., 2018  <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6222024/> | N=120  2 centers | RTx or RP | T2w, ADC | 600 intensity,  Texture features | Direct concatenation |
| Dinis Fernandes et al., 2018  <https://phiro.science/article/S2405-6316(17)30085-4/fulltext> | N=120  1 center | RTx | T2w | 254 texture features | NA |
| Zhong et al., 2020  <https://www.frontiersin.org/articles/10.3389/fonc.2020.00731/full> | N=91  1 center | RTx | T1, T2, DWI | 1536 deep features | Direct concatenation |
| Bourbonne et al., 2020  <https://www.mdpi.com/2072-6694/12/4/814> | N=195  2 centers | RP | T2w, ADC | 27,376 intensity,  Shape features | Direct concatenation |
| Li et al., 2020  <https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(20)30539-9/fulltext> | N=198  4 centers | RP | T2w, ADC | 200 intensity,  Shape features | Direct concatenation |

* Hypothesis
  + Fused radiomic features from T2w and ADC are more prognostic of BCR than:
    - 1 ) T2w radiomic features alone
    - 2) ADC radiomic features alone
    - 3) Direct feature combination from T2w and ADC
    - 4) Clinical assessments of T 2 score, T 2 shape, T 2 signal, average ADC values
* Dataset and patient selection



* + Median follow up for all 343 cases:
    - 27.9 (0.2-110.1) months
  + Median follow up for 70 BCR cases:
    - 12.1 (0.-92.9) months
  + Median follow up for 273 non-BCR cases:
    - 31.4 (1.9-110.1) months
* Radiomics feature extraction

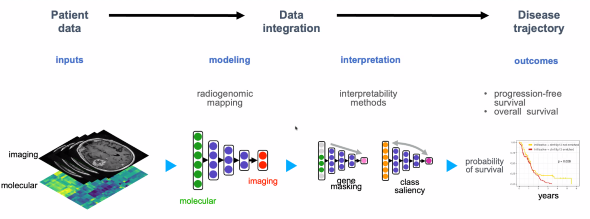




* Results
  + CCA-fused features outperform clinical, radiomic T2, Radiomic ADC, and Concatenate radiomic T2 and ADC
  + On test set, able to separate high risks and low risks groups

**Example: relating imaging to gene expression**

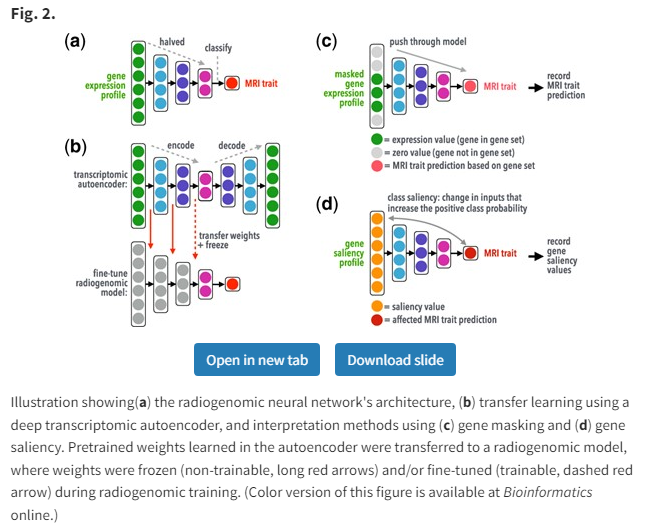
* <https://academic.oup.com/bioinformatics/article/36/11/3537/5758261>



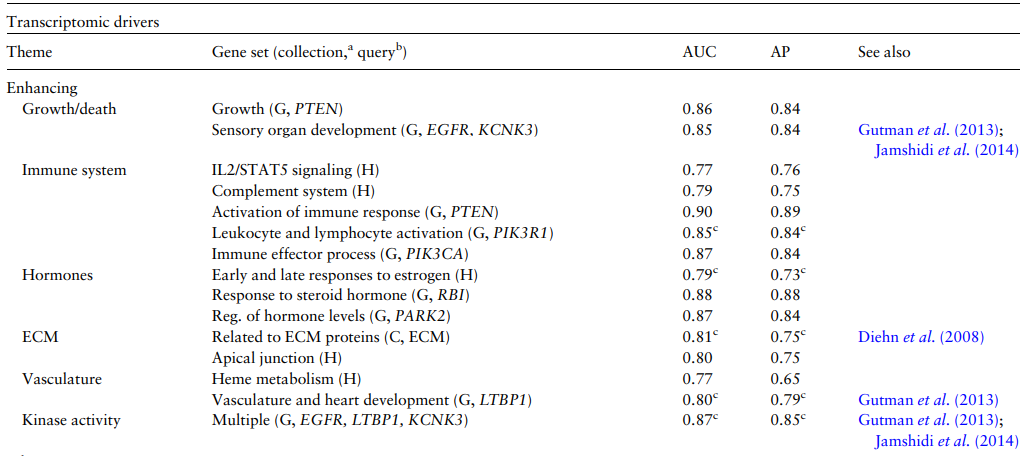
* Data:



* Radiogenomic neural network:



* Measures performance of model when only using genes from a gene set
* Use AUC as strength of association
* Attach weights to genes based on their importance in predicting a class label
* Use of GSEA to determine what gene sets (pathways, biological processes etc.) were at the top of ranked genes
* Model interpretation

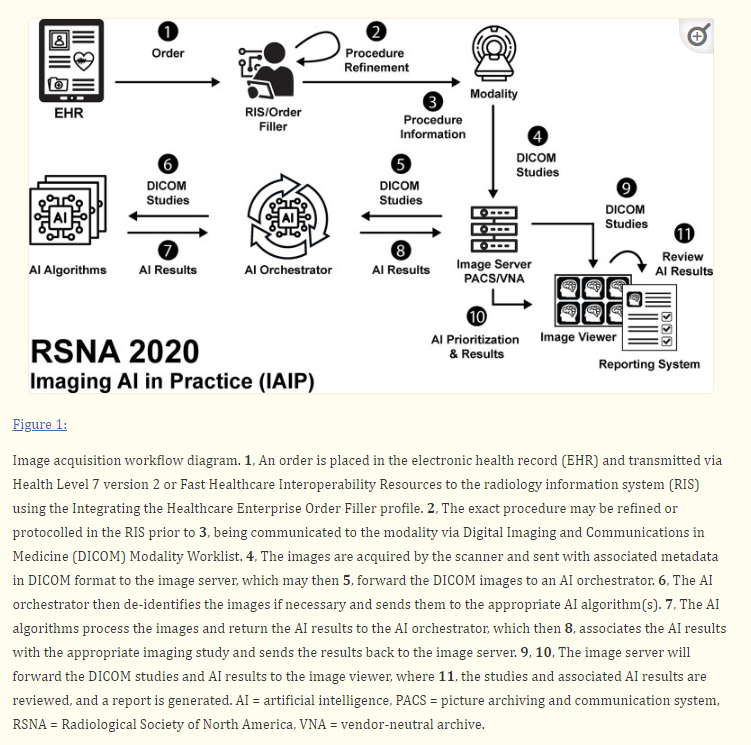


* Identified the most predictive gene sets
* Gene sets are related to growth, vasculature, immune system processes, and involved EGFR
* Gene sets align with previously identified associations

**Knowledge to practice**

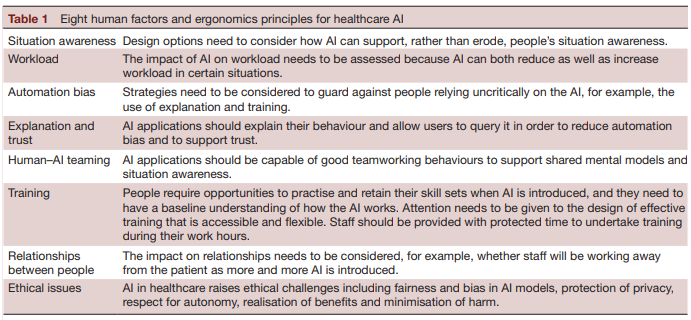


**Algorithm results are only useful if they are accessible**

****

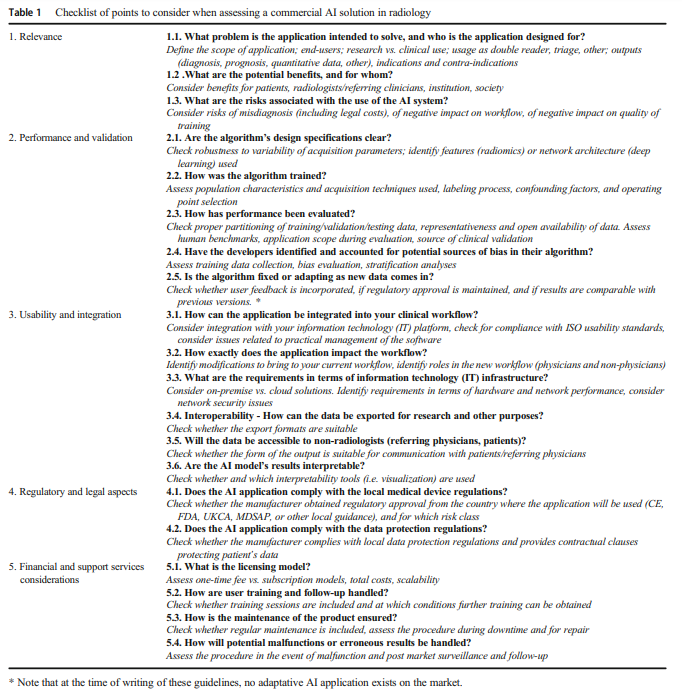
* <https://pubs.rsna.org/doi/epdf/10.1148/ryai.2021210152>

**Interaction between algorithm and human factors**

****

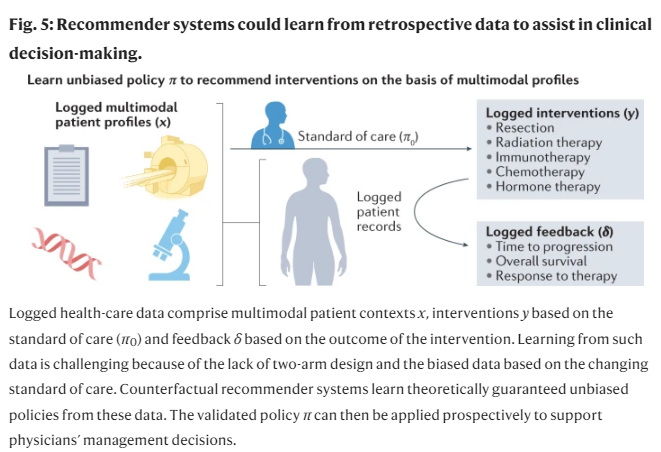
* <https://doi.org/10.1136%2Fbmjhci-2021-100516>

**Holistic view of an algorithm**



* <https://link.springer.com/article/10.1007/s00330-020-07684-x>

**Generating new data for future improvement**

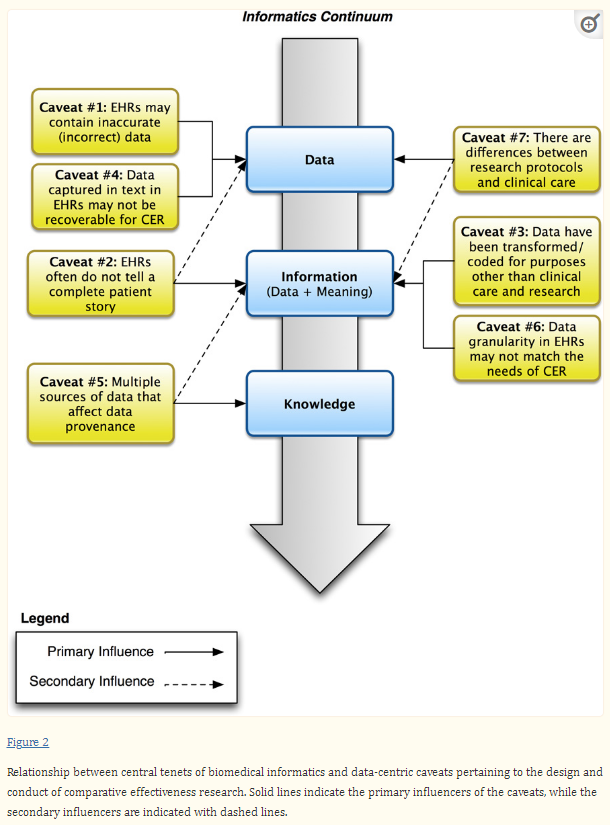


* + Boehm KM et al.
    - [https://dx.doi.org/10.1038/s41568-021-00408-3](https://dx.doi.orq/10.1038/s41568-021-00408-3)

**Getting started: datasets**

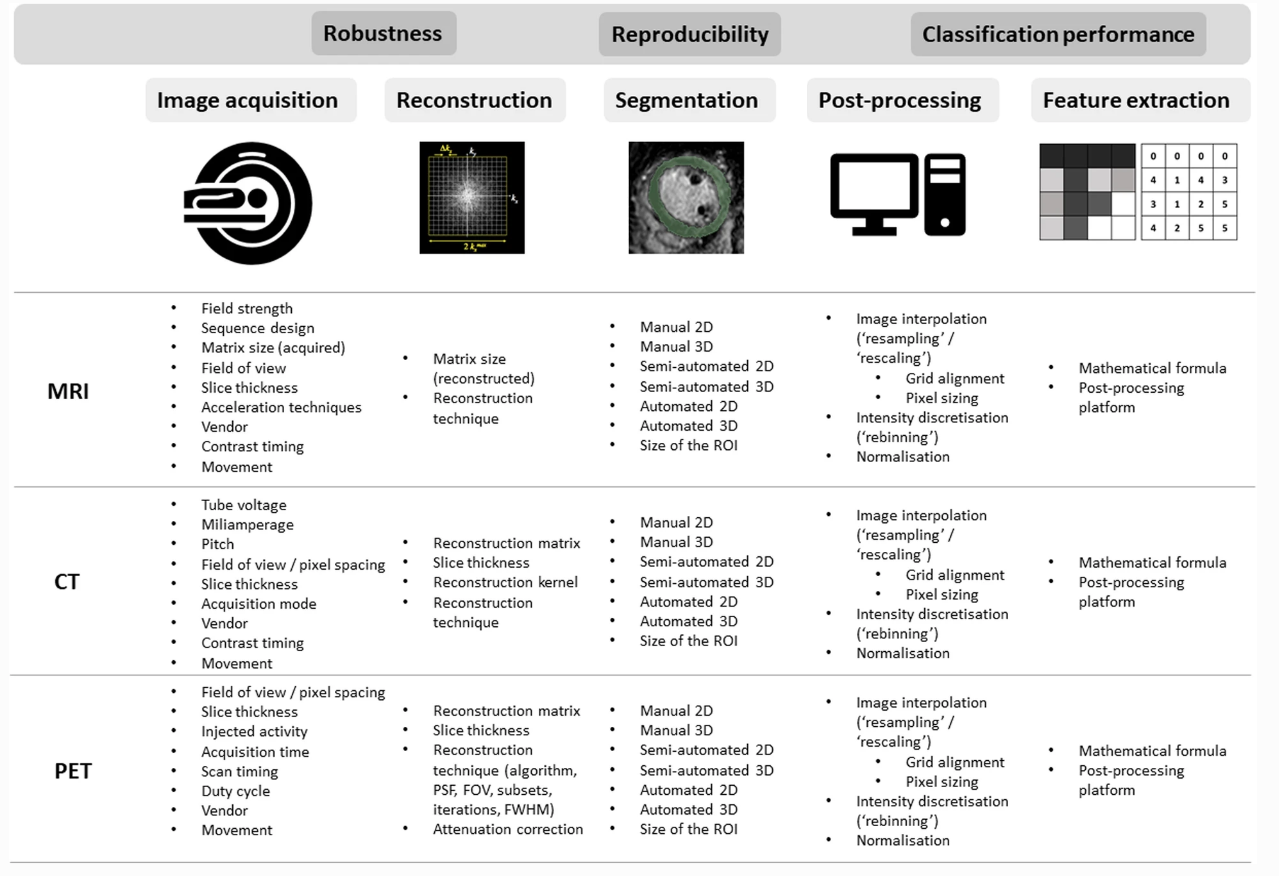
* IVY GAP Atlas Project
  + <https://glioblastoma.alleninstitute.org/>
* Cancer Research Data Commons
  + [https://datacommons.cancer.gov/](https://dataccmmons.cancer.gov/)
* Alzheimer's Disease Neuroimaging Initiative
  + <https://adni.loni.usc.edu/>
* UCLA Integrated Diagnostics Shared Resource
  + <https://idx.medsch.ucla.edu/>

**Caveats**

****

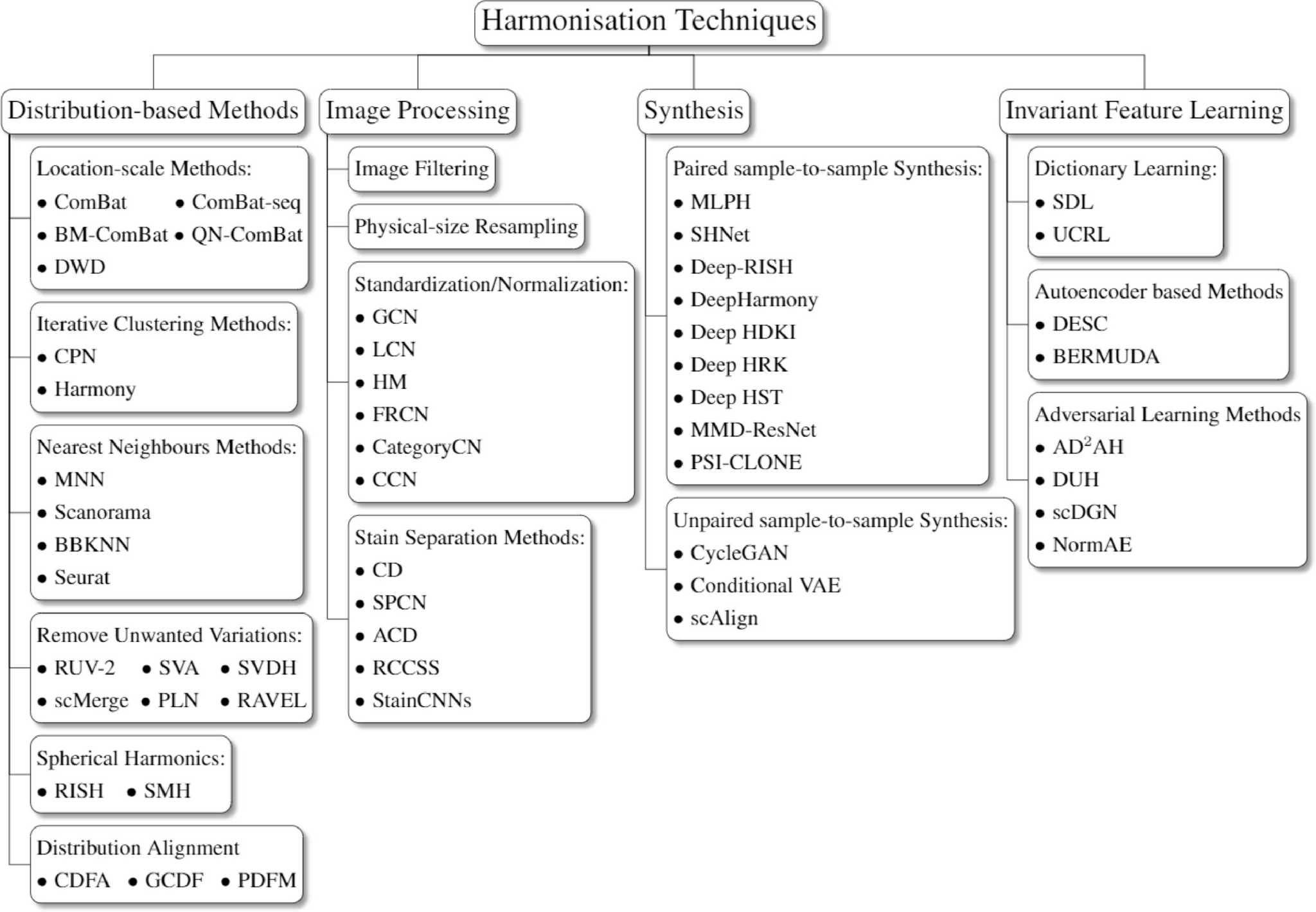
* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3748381/>

**Factors that influence radiomics stability**



* <https://insightsimaging.springeropen.com/articles/10.1186/s13244-020-00887-2>

**Methods for normalization**

****

* <https://www.sciencedirect.com/science/article/pii/S156625352200015X>

**Takeaways**

* Biomedical informatics encompasses the entire cycle from data to knowledge to practice to new data
* Machine learning, complemented by frameworks for model training, rigorous validation, and deployment with human factors, is critical to effectively harness data
* The case for multimodal data fusion
* Different modalities carry complementary (but sometimes conflicting) information
* Incorporating information across biological scales and modalities can lead to improved predictions of prognosis and treatment response
* Challenges and opportunities
* Availability of multimodal datasets
* Further investigation into how multi-modal features relate, model interpretability