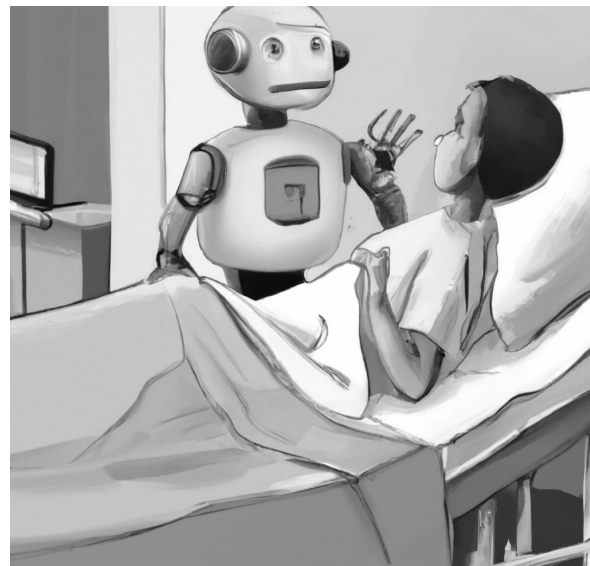


AI in medicine: a critical view on current trends

A.Chincarini
INFN / University of Genoa, Italy

andrea.chincarini@ge.infn.it



Which paradigm for future diagnostics? the current context

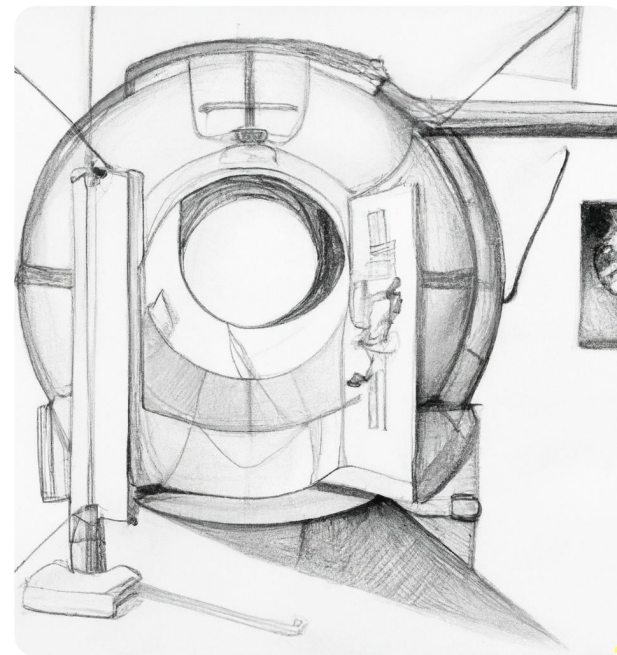
New acquisition technologies: ultra-high efficiency MRI/PET/SPECT,
multi-technology integration, software-driven spatial resolution

Highly automated and sensitive equipment will raise the number of
scans/time unit

High throughput of information

Data analysis and patient management/treatment are closely related

Data management & GDPR application is becoming more and more
taxing

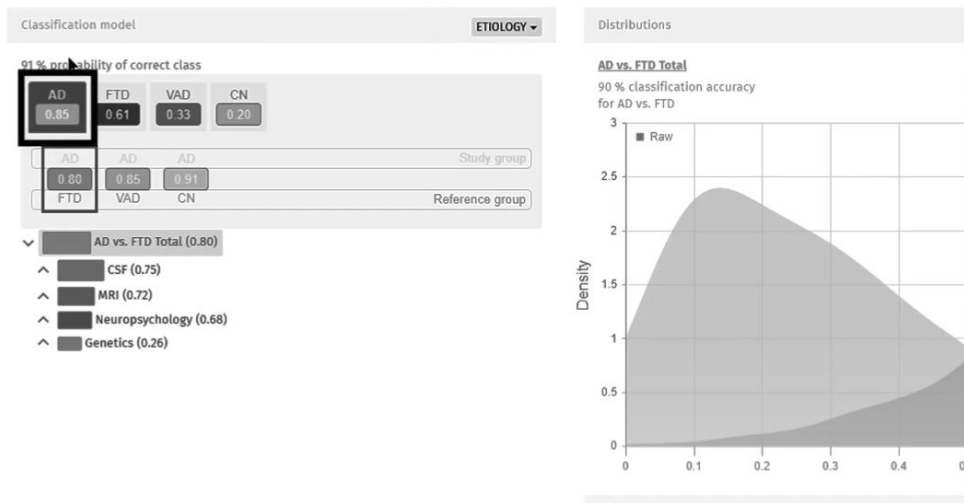


a future dense with information

information flow x 10 in the last 5 years

- the pressure to use automation is not only scientific, but also peer /marketing-induced
 - better diagnoses / fewer errors / shorter acquisition time /

are we really looking at a future where a probabilistic model will replace the clinician?



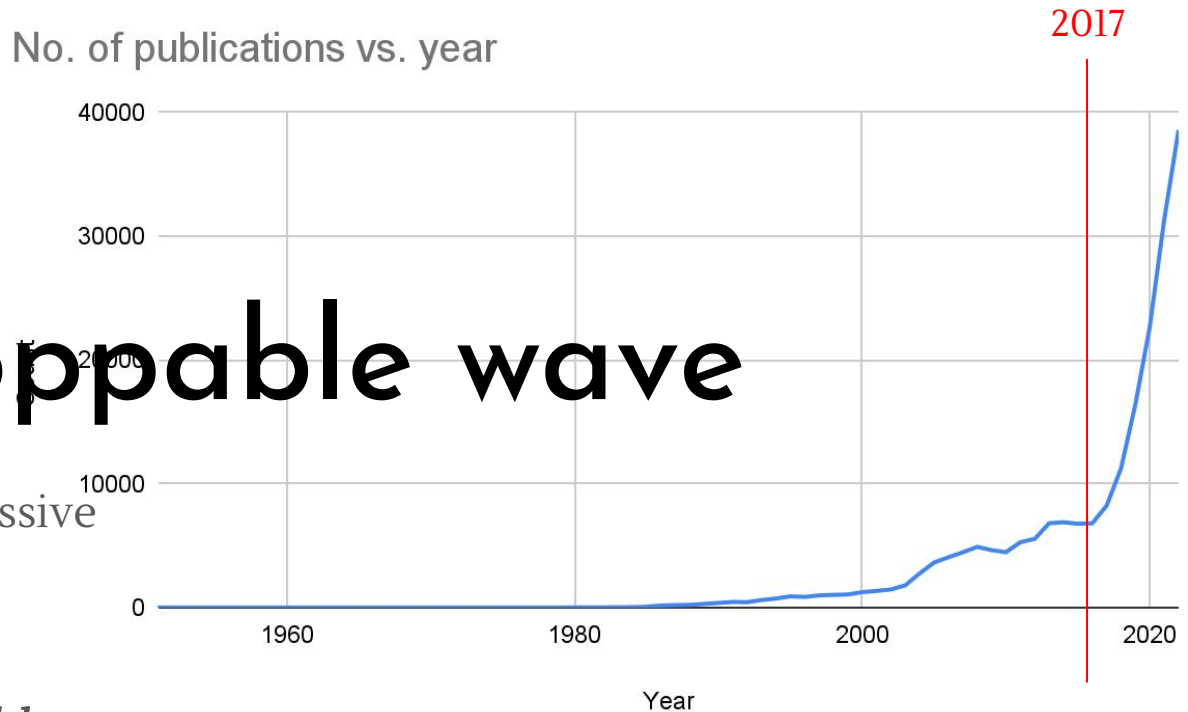
example of a commercial software in neurodegenerative disease diagnosis, whose output is a probability of belonging to a specific clinical class

the unstoppable wave

media, publications... all hint to an inevitable and massive use of AI in medicine.

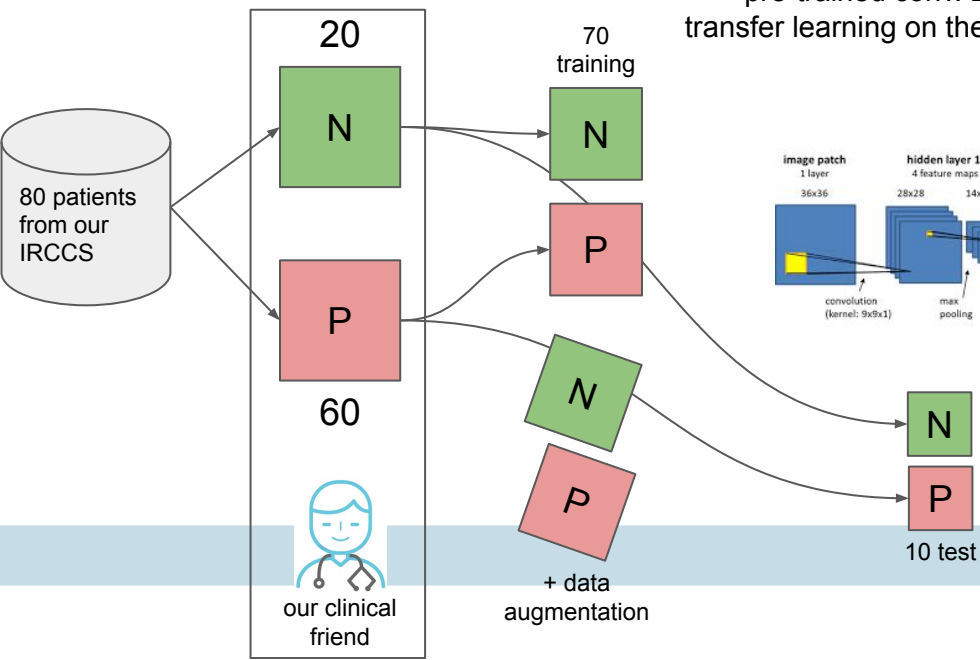
Fine.

But *all that glitters is not gold...*

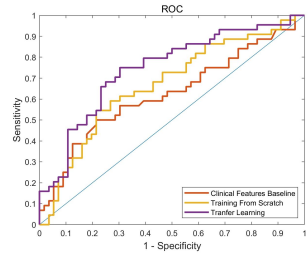
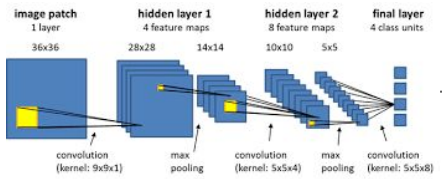


the AI frenzy is actually quite recent

a typical workflow



pre-trained conv. DL architecture*
transfer learning on the last full-conn. layer



excellent results
(cross val)



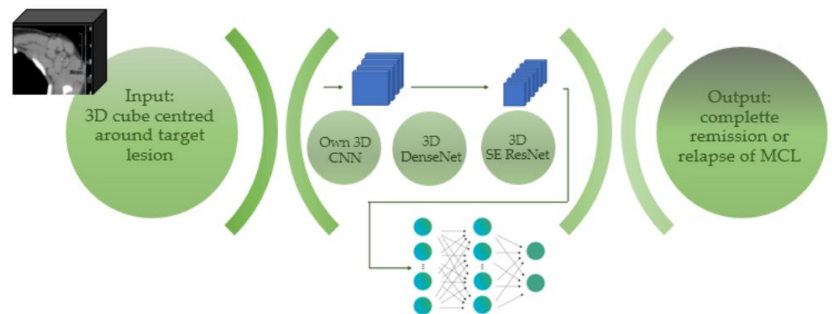
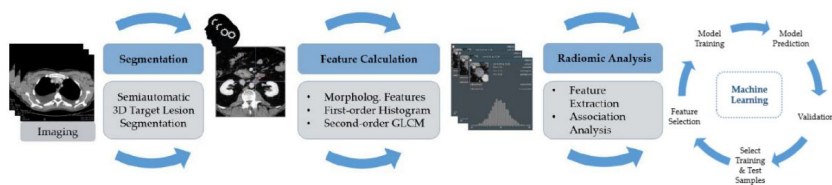
*it is not uncommon to see networks as deep as 1000 layers for analyzing <100 data
<https://github.com/KaimingHe/resnet-1k-layers>

Published online 2020 Dec 1. doi: [10.1016/j.heliyon.2020.e05625](https://doi.org/10.1016/j.heliyon.2020.e05625)

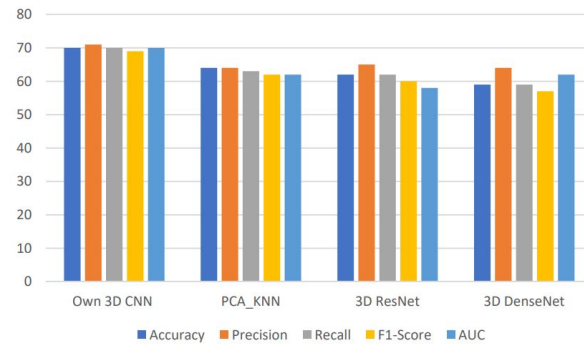
Detecting pathological brain via ResNet and randomized neural networks

Siyuan Lu,^a Shui-Hua Wang,^{a,b,**} and Yu-Dong Zhang^{a,b,*}

Article
Deep Neural Networks and Machine Learning Radiomics Modelling for Prediction of Relapse in Mantle Cell Lymphoma



Our study compared the potential of radiomics-based machine learning and 3D deep learning models as non-invasive biomarkers to risk-stratify Mantle Cell Lymphoma patients, thus promoting precision imaging in clinical oncology



Characteristic	Number
Sex	
Male	26 (86.7%)
Female	4 (13.3%)
Average age (range)	62.2 ± 9.7 years (42–76)
Ann Arbor Stage	
Stage I	0
Stage II	2 (6.7%)
Stage III	5 (16.7%)
Stage IV	23 (76.7%)
Patients' status in 5-years follow up	
Complete remission (CR)	17 (57%)
Relapse of disease (RD)	13 (43%)

a non-exhaustive
list of things that
can go wrong here
[or are definitely
wrong] but which
are unfortunately
become common
practice ...

- Fundamental & dataset
 - what is the intrinsic dimensionality of the data?
 - clinical labels by our clinical friend are taken as “true label”
 - dataset is unbalanced
 - dataset is single-center
 - dataset is “small”: are our N / P classes good representative of the “true” N / P population?
 - available metadata and clinical models (i.e. clinical knowledge) is not used
- Methodological
 - classical data augmentation (translation, rotation, inversion, stretch, ...) is typical and appropriate for image recognition processes but it has unproven advantages for other specific tasks
 - preprocessing is often neglected (on the basis that DNN/CNN do not need it / some argue preprocessing might even “damage” pre-trained DNN performance)
 - images are constrained to conform to DNN input size (because of the transfer learning)
 - DNN is trained “on cats and dogs” and not optimized onto the specific information + cross validation only on single site + sample size too small to represent the whole population = generalization is uncertain [surely bad]
 - results are taken at face value (AUC... etc). Robustness (i.e. bootstrap with balance, etc) is often neglected (too much computational cost).
 - Test on known cases (to prove the analysis is sound) or in a non-binary setting is often neglected [i.e. substitute CTRL with different pathology].
- Other
 - Explainability is often neglected.
 - publication reviewers often gaslighted by use of AI... very few have the competence to dive deeper ⇒ tons of publications with incredible results but very low reliability
 - even if all is fine, the application would assume that the patient in input is either N or P. there is no other possible outcome

part 1

fundamental issues related to
medical data, sampling &
ground truth

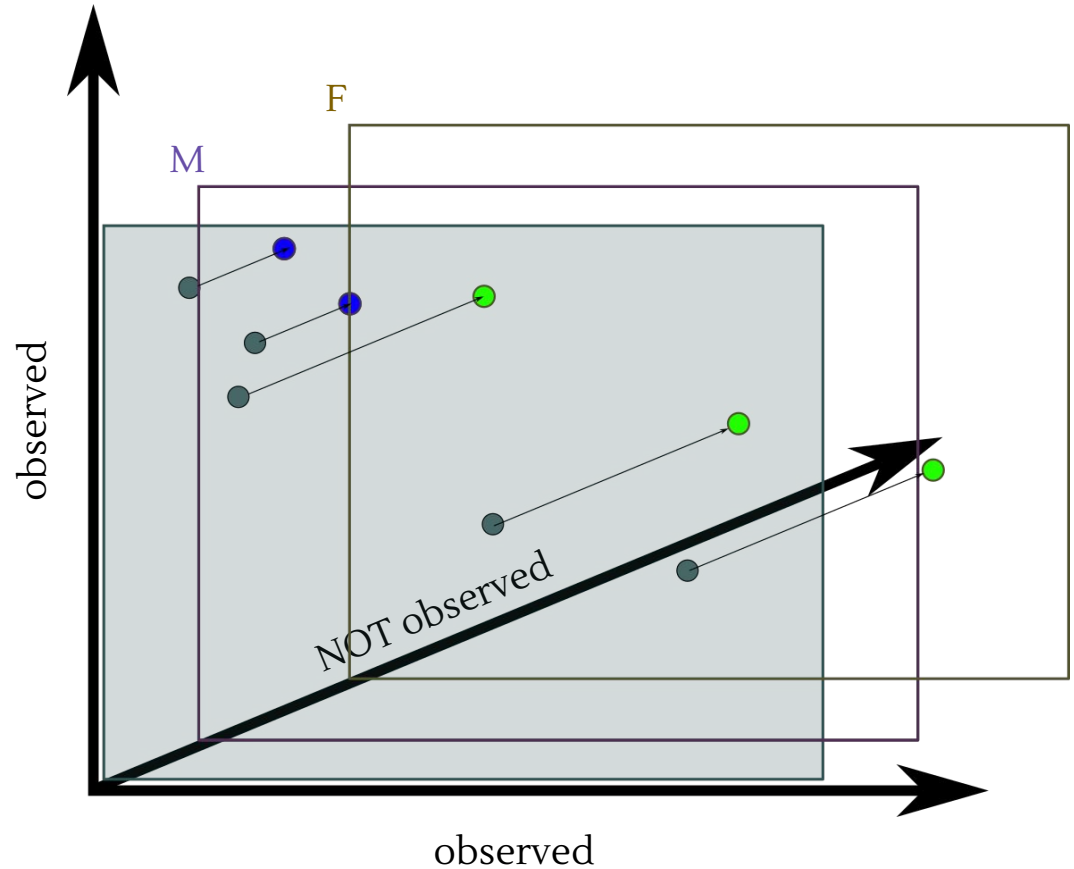
- the intrinsic dimensionality of the data is much higher than the observed variables
- clinical labels by our clinical friend are taken as “true label”
- dataset is unbalanced
- dataset is single-center
- dataset is “small”: are our N / P classes good representative of the “true” N / P population?
- ...

the “unobserved dimensions” problem

we often do not know the actual dimensions in which our data live

there are several unobserved variables with relevant implication in the data (if they were observed)

rules learned on the dataset are not trustworthy



the “true label” problem

a couple of examples from:

amyloid PET

lung CT (COVID cases)

typically:

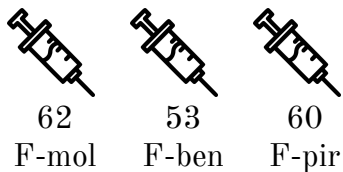
- dataset are never evaluated by more than one human expert
- approx. ground truth can sometimes be achieved by lab examinations ...
 - i.e. genetics / ...
- ... but almost never on clinical assessment
 - i.e. diagnosis / segmentations / ...
- can we measure uncertainty on a clinical diagnosis?

testing evaluation #1: the experiment



5 independent expert readers

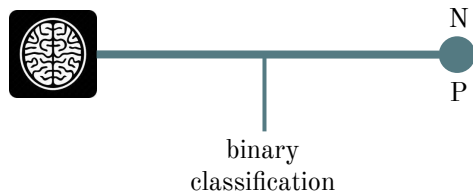
We asked 5 NM phys to evaluate 175 scans (^{18}F)
2 expert and 3 medium-expert evaluators



3 fluorinated tracers
(175 cases)

Readers never interacted, they were to label each scan independently.

We *didn't* look for consensus.



binary evaluation

Complicated dataset! 6 EU centers, clinical-grade scans, no shared protocol.

No scan was rejected.

in the end...

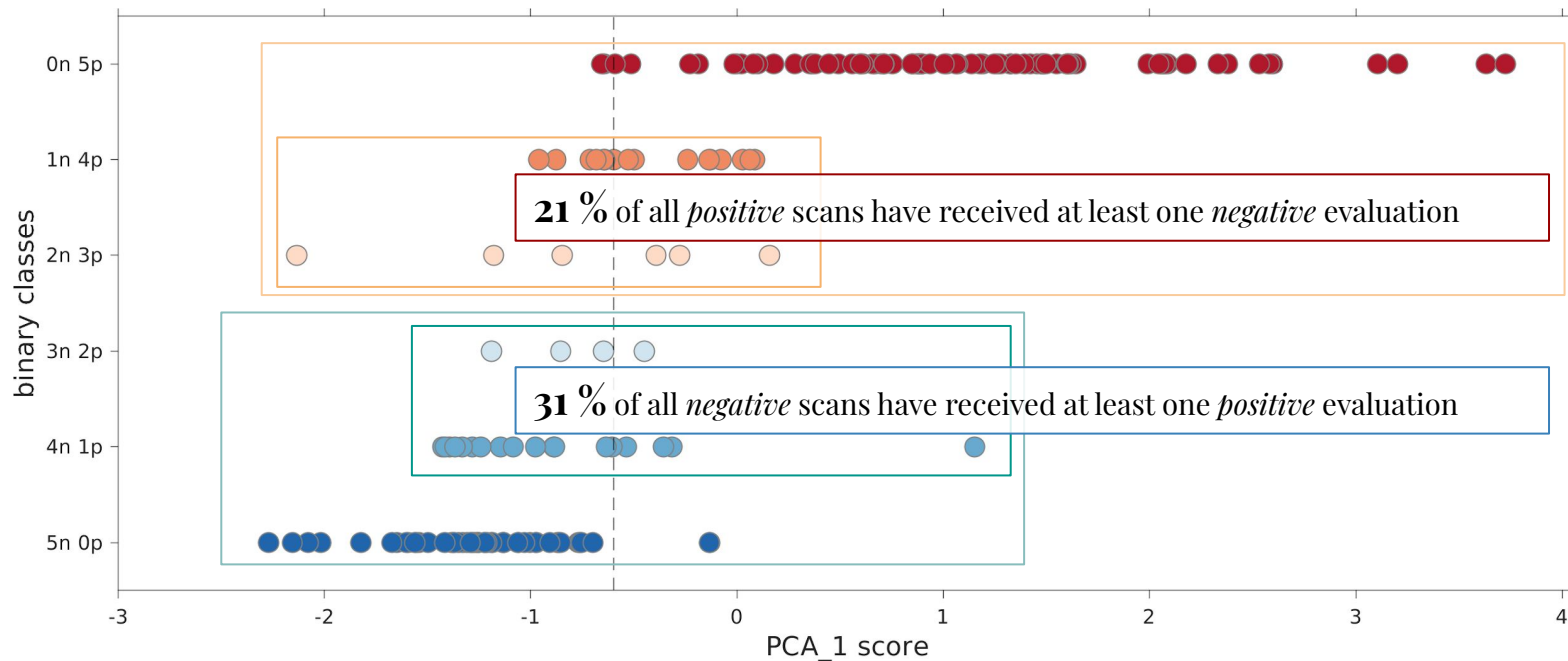
6 EADC centers: ANT, BRE, GEN
HUG, MAN, PAR

multicenter,
clinical-grade scans

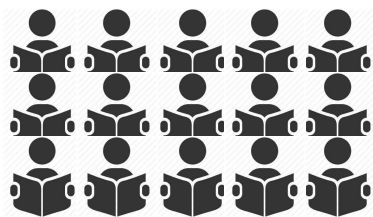
Each scan received 5 label (either **N** or **P**, one for each reader).

2 semi-quantification
(SUV_T , ELBA)

where do we go wrong?



testing evaluation #2: the experiment



14 independent
expert readers

We asked 14 radiologists to evaluate 120 lung CT scans
9 expert and 5 medium-expert evaluators

120 lung CT from
EU hospitals

Readers never interacted, they were to label each scan independently.

binary evaluation

We *didn't* look for consensus.

Complicated dataset! anonymous dataset,
clinical-grade scans, no shared protocol.

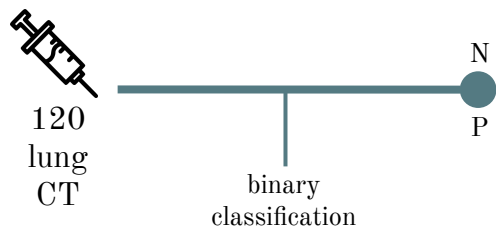
No scan was rejected.

multicenter,
clinical-grade scans

in the end...

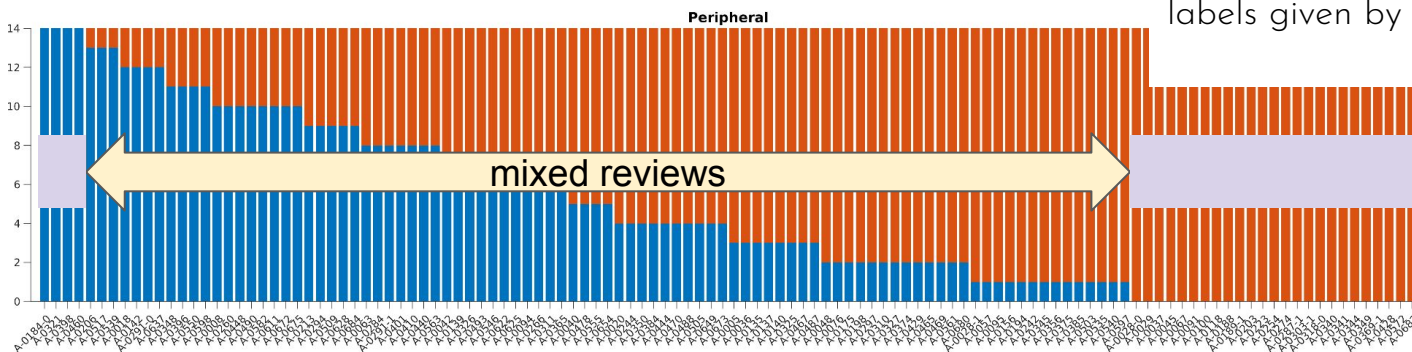
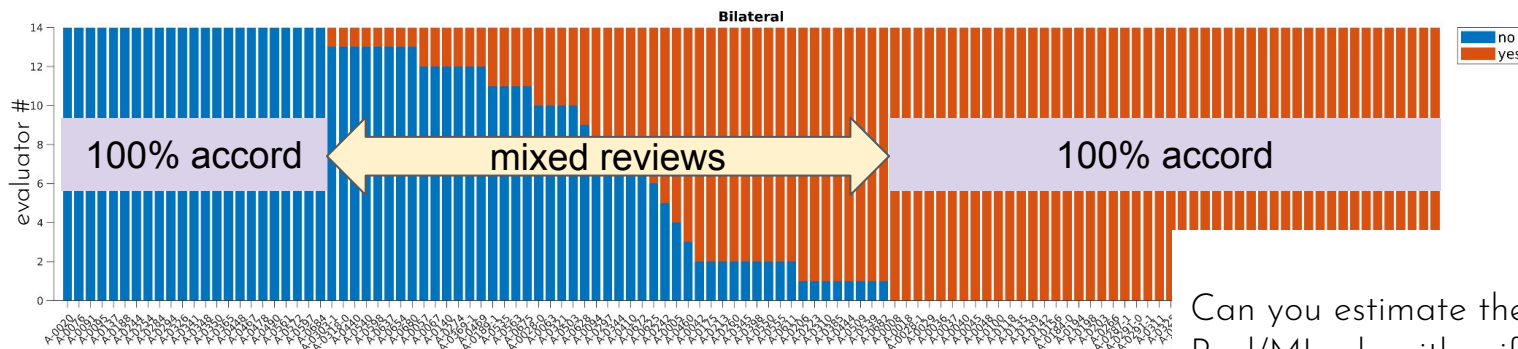
Each scan received 2 binary labels, a 0-100 grading
(R/L) and the affected volume grade assessment

1 quantification
(lungQUANT)



14 readers in 9 centers

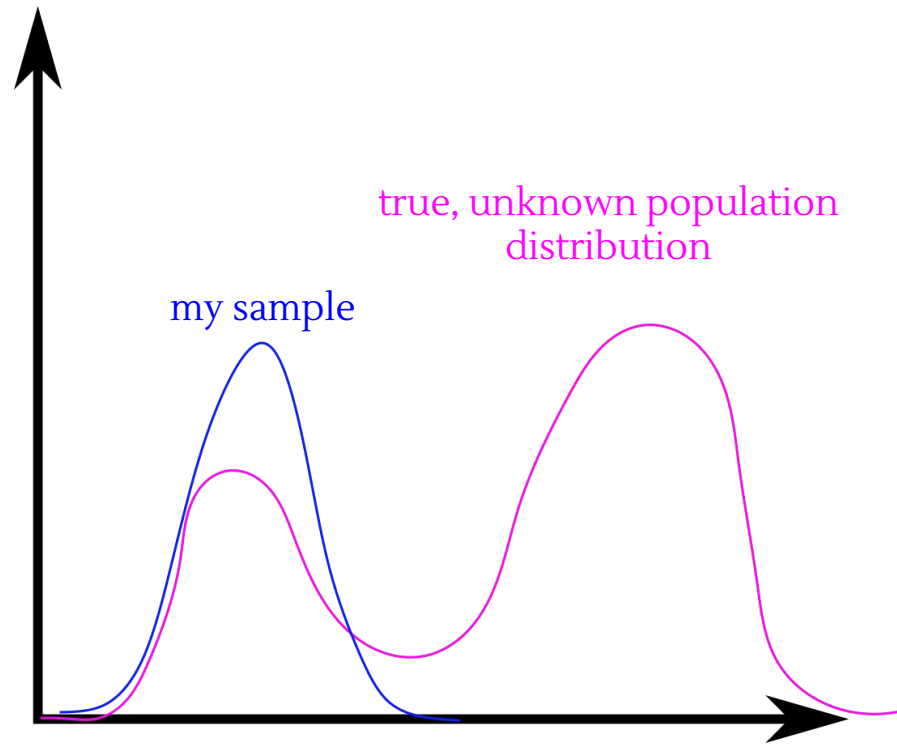
Evaluation heterogeneity by case



Can you estimate the error in your Rad/ML algorithm if you trained on labels given by a single clinician?

the “sampling bias” problem

- no further comment needed ...
- this situation is more common than people would have you believe
 - an EADC study on the topic showed impressive difference in in-patient statistics both among nations and within centers
- a relatively simple bias which is difficult to correct
 - large multicentric studies should do the trick but at the cost of increasing the batch (provenance) effect



part 2

methodological issues related to generalization and data-specific characteristics

- classical data augmentation (translation, rotation, inversion, stretch, ...) is typical and appropriate for image recognition processes but it has unproven advantages for other specific tasks
- preprocessing is often neglected (on the basis that DNN/CNN do not need it / some argue preprocessing might even “damage” pre-trained DNN performance)
- images are constrained to conform to DNN input size (because of the transfer learning)
- DNN is trained “on cats and dogs” and not optimized onto the specific information + cross validation only on single site + sample size too small to represent the whole population = generalization is uncertain [surely bad]
- results are taken at face value (AUC... etc). Robustness (i.e. bootstrap with balance, etc) is often neglected (too much computational cost).
- Test on known cases (to prove the analysis is sound) or in a non-binary setting is often neglected [i.e. substitute CTRL with different pathology].

on data augmentation

augmentation is perfectly adequate and proper with image recognition tasks

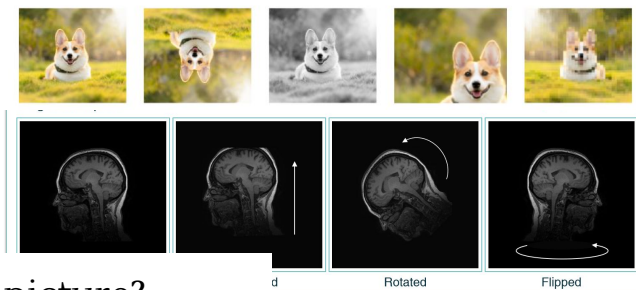
augmentation is **beneficial** when the **added variability spans the same domain** as the information (signal) we want to target

however

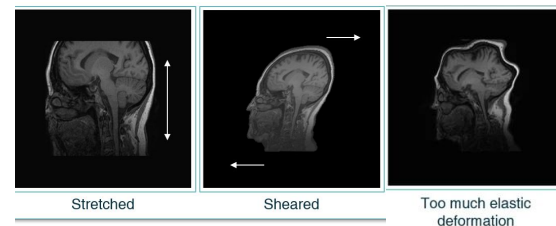
general augmentation techniques are **detrimental** when they add variability in a domain unrelated to the signal (i.e. it only adds confounding information)

one is lead to believe that feeding the CNN with raw data is better than focusing the discriminating power onto the actual signal... wrong!

why not feed the CNN with a spatially normalized scan?



is there a dog in the picture?
is there a human head in the MRI?



is there atrophy in the hippocampus?

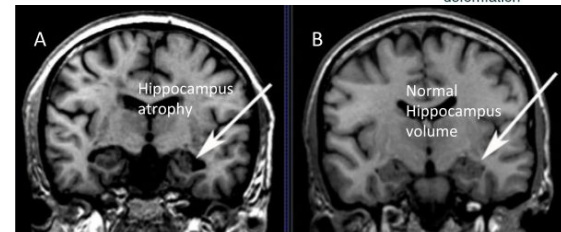


Figure 1: A = Brain MRI from a patient with Alzheimer disease, showing reduction of the hippocampus's volume. B = Brain MRI from a no Alzheimer patient showing normal volume of hippocampus.

transfer learning: is it truly THE solution?

data/images are constrained to conform to DNN input size

DNN is trained on much different data than medical information / assume feature space is large enough to span all situations but signal in medicine is much subtler than the one in the original dataset

Soliman et al.
BMC Medical Informatics and Decision Making (2022) 22:318
<https://doi.org/10.1186/s12911-022-02054-7>

BMC Medical Informatics and
Decision Making

RESEARCH

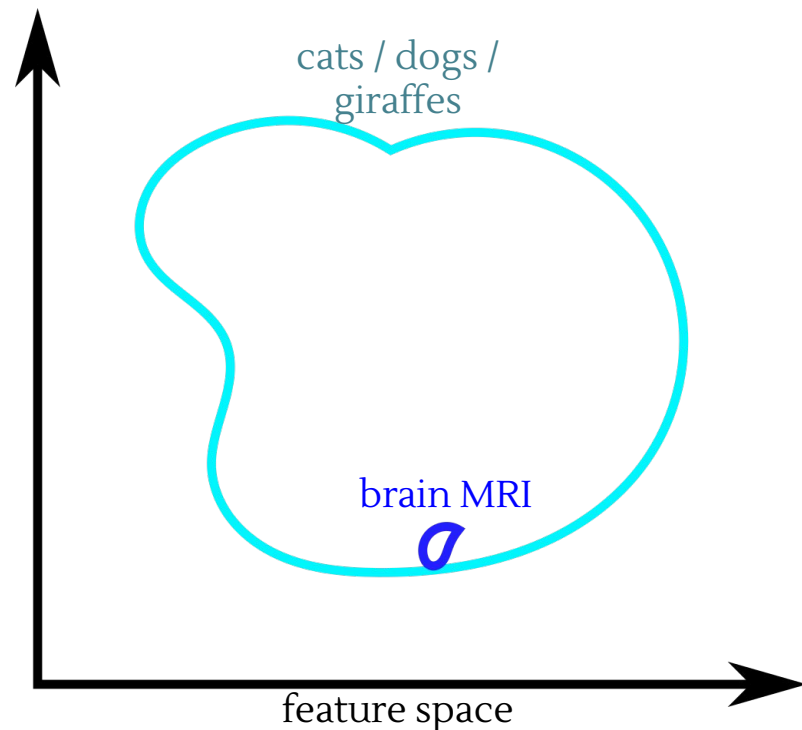
Open Access



Adopting transfer learning for neuroimaging: a comparative analysis with a custom 3D convolution neural network model

Amira Soliman^{1*}, Jose R. Chang^{1,2}, Kobra Etrminani¹, Stefan Byttner¹, Anette Davidsson¹, Begoña Martínez-Sánchez⁴, Valle Camacho⁵, Matteo Bauckneht⁶, Roxana Stegeran⁷, Marcus Ressler⁸, Marc Agudelo-Cifuentes⁴, Andrea Chincarini⁹, Matthias Brendel¹⁰, Axel Rominger¹¹, Rose Bruffaerts¹², Rik Vandenbergher^{12,13}, Milica G. Kramberger¹⁴, Maja Trost¹⁵, Nicolas Nicastro¹⁶, Giovanni B. Frisoni¹⁷, Afina W. Lemstra¹⁸, Bart N. M. van Berckel¹⁹, Andrea Pilotto²⁰, Alessandro Padovani²¹, Silvia Morbelli²², Dag Aarsland^{23,24}, Flavio Nobili²⁵, Valentina Garibotto²⁶, the Alzheimer's Disease Neuroimaging Initiative and Miguel Ochoa-Figueroa^{3,25}

“... additionally, custom 3D model performs comparably to TL models for binary classification, and interestingly perform better for diagnosis of multiple disorders. The results confirm the superiority of the custom 3D-CNN in providing better explainable model compared to TL adopted ones.”



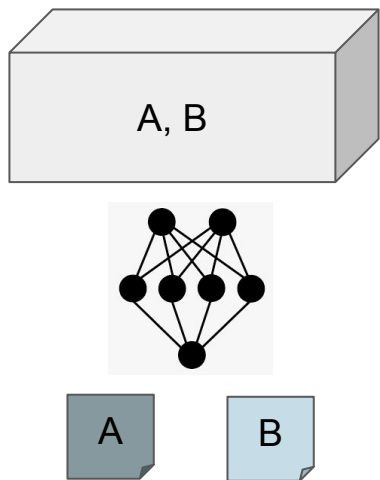
part 3

application and “social” issues,
AI expectations

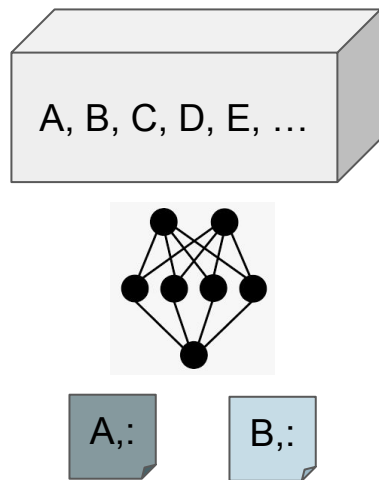
- partition bias: the algorithm assumes that the data in input is either N or P. No other outcome is contemplated
- available metadata and clinical models (i.e. clinical knowledge) is not used
- explainability is often neglected.
- publication reviewers often gaslighted by use of AI... very few have the competence to dive deeper \Rightarrow tons of publications with incredible results but very low reliability
- there is a psychological bias linking beliefs to complex, AI methodologies

partition bias

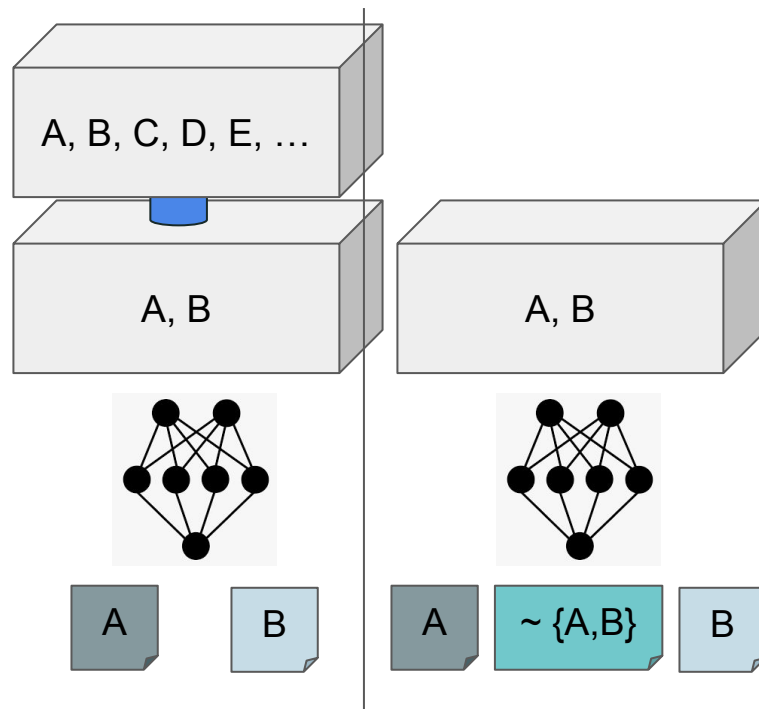
what we are doing



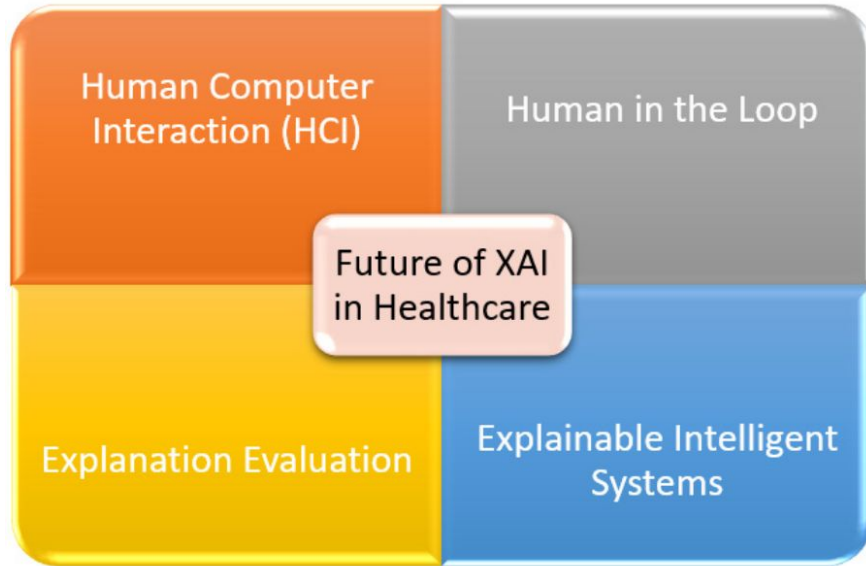
what happens in application



what we should
[probably] do ...



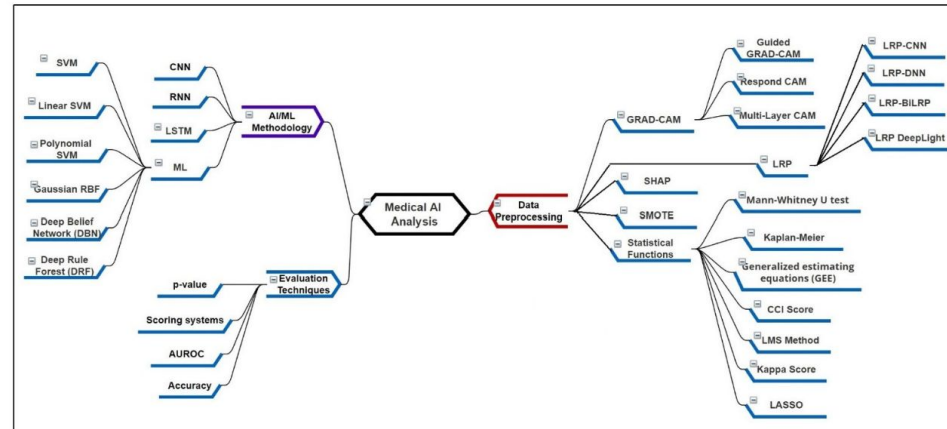
XAI and the importance of knowledge embedding



Review

A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System

Ruey-Kai Sheu ¹ and Mayuresh Sunil Pardeshi ^{2,*}



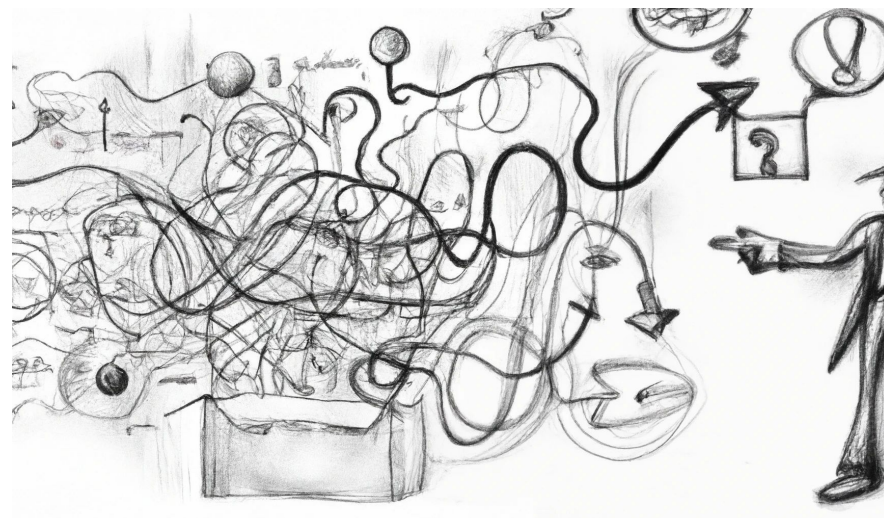
magical thinking (psychological bias)

- people tend to lower their critical thinking when AI is involved:
 - because of the perceived intelligence of the machines.
 - it is assumed that the AI is able to process and analyze data in a way that is beyond human capabilities

Artificial Intelligence and Magical Thinking

✍ CHARLES CORFIELD, PRESIDENT & CEO, NVOQ / 📅 MAY 14, 2019 / 🏢 HEALTHCARE / 💬 LEAVE A COMMENT


lack of skepticism and critical examination of the output.



- Hype and marketing of AI, the promise of the AI to make things easier, faster and more accurate:
 - belief that the AI is capable of solving all the problems.

lack of critical thinking and a belief that the AI is always right.

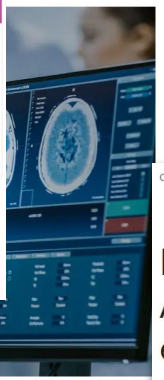
CASE STUDY
**Self-Documenting ML Experiments
Slash Wasted Time for Philips**
NOVEMBER 10, 2020



Magic AI
Ospedali e strutture sanitarie

Segui


Chi siamo
Magic-AI is a medical networking platform created to harnesses the collective ingenuity of clinicians combined with the power of deep learning technologies throughout the consultation and patient management process.




The Solution

The Philips team's search brought them to ClearML, which ticked all module, the system began tracking and managing data experiment: their algorithm team. Immediately, they felt the difference as efficie

First, ClearML's experiment manager "automagically" captures metr the research basically documents itself. "It was impressive to watch Algorithms Group Leader. "It was like we'd hired a whole team of pe



Journal of Biomedical Informatics
Volume 84, August 2018, Pages 184-199



From narrative descriptions to MedDRA:
automagically encoding adverse drug reactions

CSU News > Opinion > Exciting magic, or misanthropic sorcery? Artificial intelligence and medical imaging explained

**Exciting magic, or misanthropic sorcery?
Artificial intelligence and medical imaging
explained**

DESIGNLINES | AI & BIG DATA DESIGNLINE

6 APRIL 2022

**When We Put AI in Medical Devices,
Magic Starts to Happen**

By Sally Ward-Foxton 07.12.2022

Share Post [Share on Facebook](#) [Share on Twitter](#) [in](#)



trusting the algorithm

no need to explain what sometimes happens when one follows software-suggested routes blindly.

There are several reasons why people tend to have a high level of trust in AI or complex algorithms.

1. these systems are often seen as objective and unbiased,
 - a. belief that their outputs are always accurate.
2. the complexity and technical nature of AI and algorithms can make them seem more credible and trustworthy
 - a. the technology is too advanced for us to understand.
3. success in other scientific fields
 - a. lead to analogically translate it to any application



AI: cutting through the hype

- the impressive advances in other application fields (Google, openAI, etc)
 - suggest that the same can happen in any field
 - spun the belief in the golden age of AI
 - fosters the blind trust in what comes out

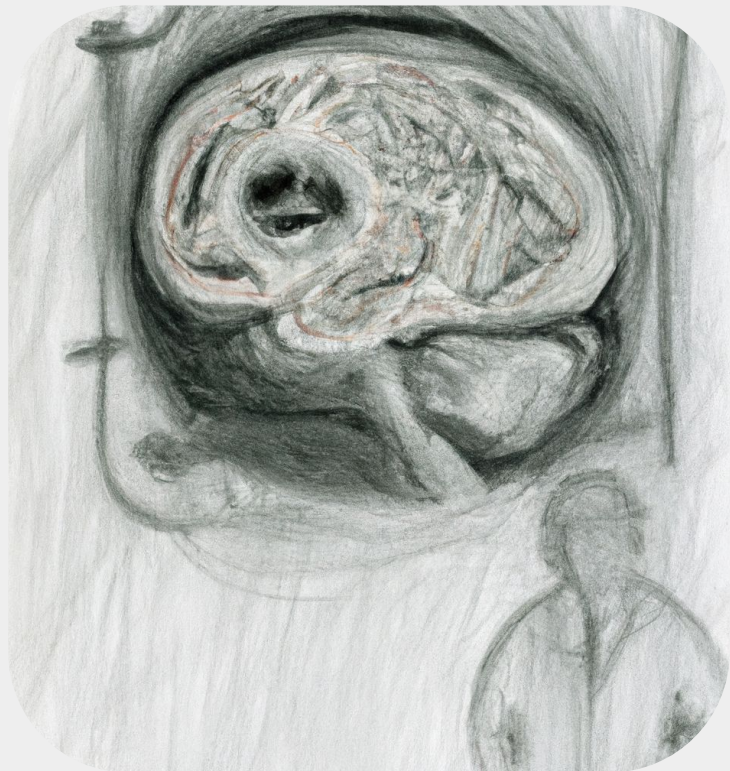
Recent studies discuss the limitation and the potential harm linked to the overreliance on AI-related tools in medicine, but they are too few ...

gender/age/race bias, risk to the patient, damage to the diagnostic process, legal matters, ...

- Fosso Wamba, S. (2022). Impact of artificial intelligence assimilation on firm performance: The mediating effects of organizational agility and customer agility. *International Journal of Information Management*, 67, 102544. <https://doi.org/10.1016/j.ijinfomgt.2022.102544>
- Adam, H., Balagopalan, A., Alsentzer, E., Christia, F., & Chassei, M. (2022). Mitigating the impact of biased artificial intelligence in emergency decision-making. *Communications Medicine*, 2(1), 149. <https://doi.org/10.1038/s43856-022-00214-4>
- Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. <https://doi.org/10.1080/0960085X.2021.1927212>
- *The Ethics of Artificial Intelligence in Healthcare and Research NordForsk Event (DK) (2021)*
- Ahmad, Z., Rahim, S., Zubair, M., & Abdul-Chafar, J. (2021). Artificial intelligence (AI) in medicine, current applications and future role with special emphasis on its potential and promise in pathology: present and future impact, obstacles including costs and acceptance among pathologists, practical and philosophical considerations. *A comprehensive review. Diagnostic Pathology*, 16(1), 24. <https://doi.org/10.1186/s13000-021-01085-4>
- Budd, S., Robinson, E. C., & Kainz, B. (2021). A survey on active learning and human-in-the-loop deep learning for medical image analysis. *Medical Image Analysis*, 71, 102062. <https://doi.org/10.1016/j.media.2021.102062>
- Matheny, M. E., Whicher, D., & Thadanev Israni, S. (2020). Artificial Intelligence in Health Care. *JAMA*, 323(6), 509. <https://doi.org/10.1001/jama.2019.21579>
- Keris, M. P. (2020). Artificial intelligence in medicine creates real risk management and litigation issues. *Journal of Healthcare Risk Management*, 40(2), 21–26. <https://doi.org/10.1002/jhrm.21445>
- Jung, C. L. (2019). The perils of artificial intelligence in healthcare: Disease diagnosis and treatment. *Journal of Computational Biology and Bioinformatics Research*, 9(1), 1–6. <https://doi.org/10.5897/JCBBR2019.0122>
- Kabir, M. (2019). Does artificial intelligence (AI) constitute an opportunity or a threat to the future of medicine as we know it? *Future Healthcare Journal*, 6(3), 190–191. <https://doi.org/10.7861/fhj.teale-6-3>
- Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3), 231–237. <https://doi.org/10.1136/bmjqs-2018-008370>
- McCartney, M. (2018). Margaret McCartney: AI in medicine must be rigorously tested. *BMJ*, k1752. <https://doi.org/10.1136/bmj.k1752>
- Ranschaert, E. (2018). Artificial Intelligence in Radiology: Hype or Hope? *Journal of the Belgian Society of Radiology*, 102(S1). <https://doi.org/10.5334/jbsr.1632>

part 4

solutions?



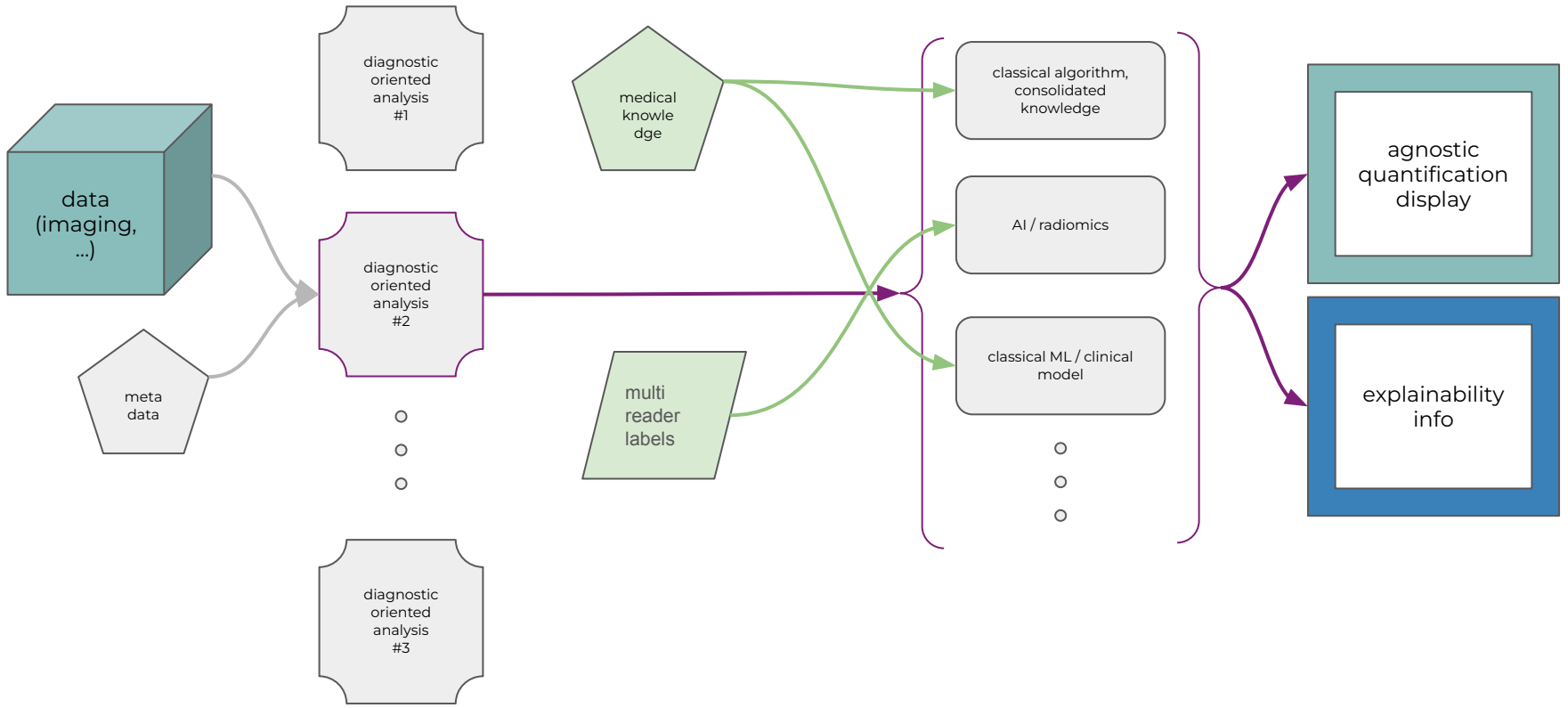
Considerations

Overall, AI and complex algorithms can be powerful tools but:

- they are not infallible and human oversight and critical thinking are still necessary to ensure accurate and trustworthy results
- no matter how sophisticated the technology is, it is still a tool created by humans and can have errors, biases, and limitations that should be constantly monitored
- *the devil is in the details...* oftentimes a seemingly rigorous procedure (in the algorithmic sense) might fail because of incomplete information / unknown biases / or simply because of compartmentalized knowledge (data scientist / clinicians / end user)
- aim for an AI that complement, not replace
 - the expertise and intuition of a qualified human
- embed quality checks in the process
 - data / intermediate results / references
- use explainability with multiple approaches
 - crucial in ensuring reliability and result communication
 - target XAI at human (i.e. end user) understanding
- ethical considerations are important in the use of AI in medicine
 - transparency in the dataset is as important as in algorithm
- sometimes it is unwise to use advanced methods
 - Know when to restrain yourself

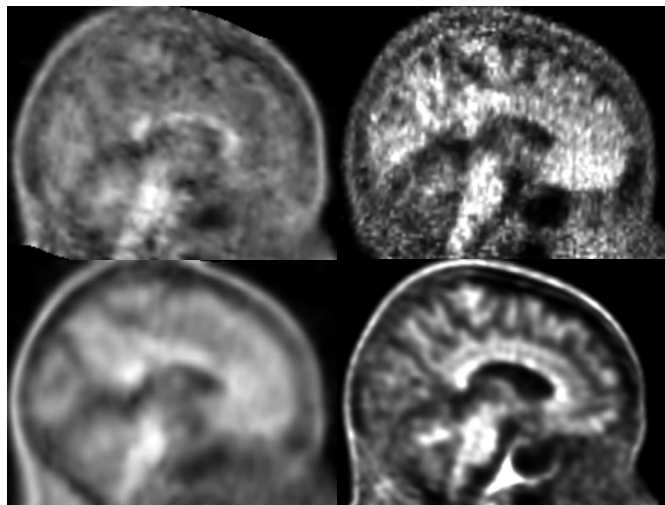
a possible framework for ethical AI in medicine

1. **include** clinicians/medical experts in the process from the beginning. Make a clear statement about the question.
2. **embed** as much knowledge as possible
3. **do not trust** AI only, but do not "throw the baby out with the bathwater": use **multiple, independent** approaches with both traditional and advanced analyses
4. keep implementation as **simple** as possible, don't go for the most complex algorithm first
5. **aim at quantification** (i.e. a direct product of the exam, open to interpretation) rather than at the diagnostic label (the outcome of a more complex process often involving several experts, difficult to challenge)
6. be **specific**: one analysis targeting a specific issue at a time (avoid general purpose approaches / the "do-it-all-end-of-the-world" algorithm). combine into higher-order analysis only after extensive validation
7. **train users** & **explain** algorithms
8. **know the limitations** of your training dataset, algorithm and application
9. implement **continuous training**: embed new info as it becomes available & check convergence on previous and new results
10. test your findings with a dimensionally **compact representation** of your data



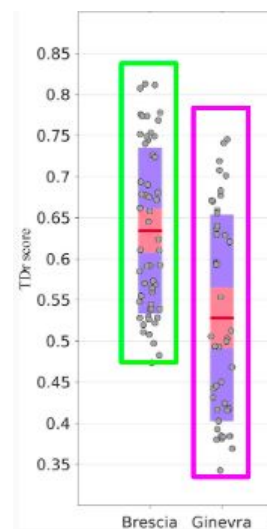
further considerations: data harmonization

small single-center dataset \Rightarrow multi-centric studies \Rightarrow provenance (batch) effect



4 samples from the EADC dataset [900 amy-PET scans]
Data after MNI spatial-normalization

scan quality
sign. affects
measures but...



center label		known demographic variables
batch effect	unknown demographic variables	

LeWinn, K.Z., et al. Sample composition alters associations between age and brain structure. *Nat Commun* 8, 874 (2017)

nature communications

“... we find that the distribution of basic socio-demographic characteristics within a study sample, including race/ethnicity and socio-economical status, **meaningfully influences** the association between age and brain structure.”

Can we decouple “quality” from “center” ?

Difference is due to: **batch** [acq. protocol, reconstruction, ...] + **demographic var.** [known] + **other variables** [unknown]

radiomic approach to batch effect mitigation

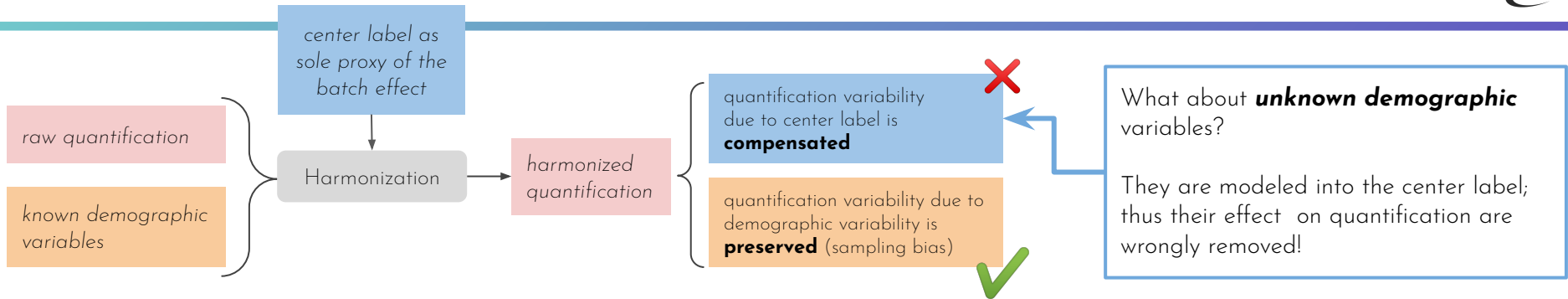
Assumptions

1. That which does not affect the image can be neglected
2. NMI quality is affected by acquisition, scanner, rec. protocol ..., but not by the clinics
3. In NMI scans one can always define 2 VOIs: one tracer-specific and one non-specific.
4. The non-specific signal does not relate to any physiological variability relevant to the pathology under consideration

Consequences

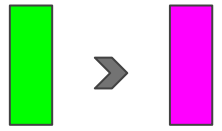
- A. We can define radiomic features to describe the heterogeneity of our dataset. This is equivalent to a *n-points scale calibration*
 - a. typical features = texture
- B. These will be applied to non-specific VOI to characterize **only** the batch effect (= image quality)
- C. After dimensionality reduction, these can be used **instead of** the center label to regress the batch effect using any suitable harmonization procedure (ComBat*, ...)

*<https://github.com/Jfortin1/ComBatHarmonization>



A-POSTERIORI HARMONIZATION

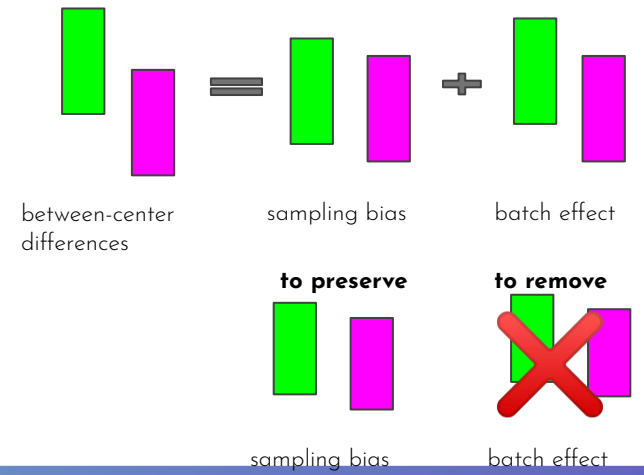
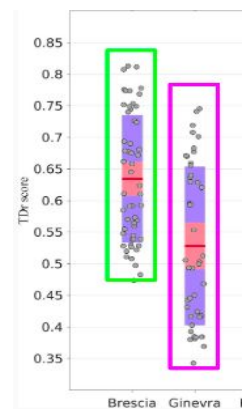
Suppose there is an unknown age bias



mean 80 yrs mean 70 yrs

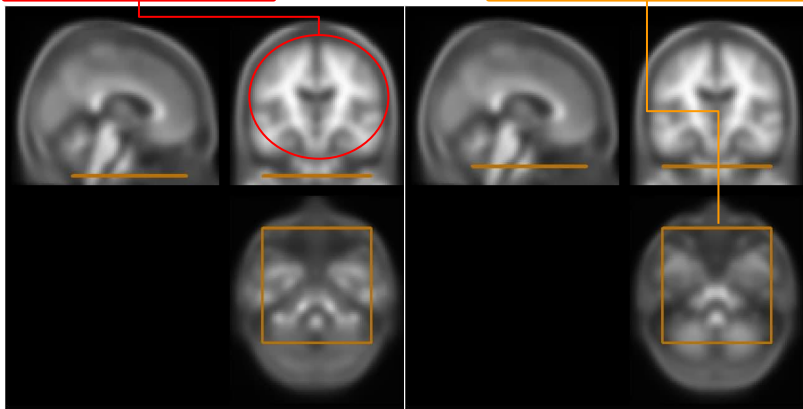
Sampling bias play an important role!

Sampling bias effects on quantification must be preserved!



specific VOI
brain parenchyma

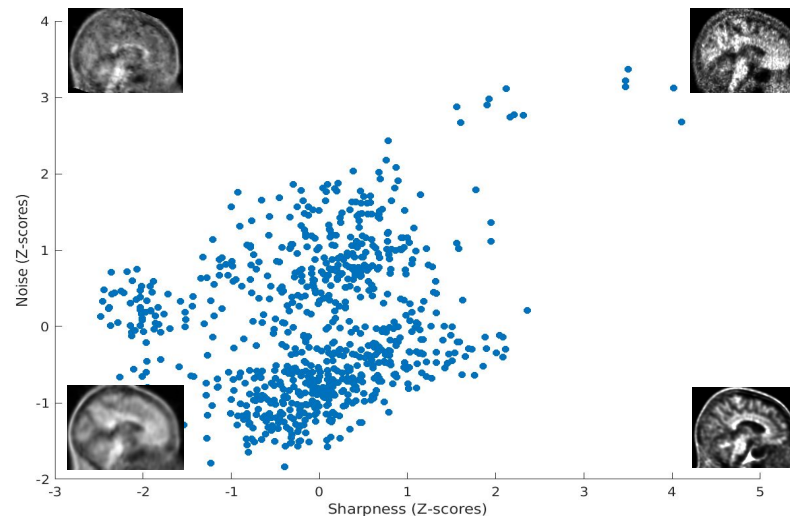
non-specific VOI
brain stem, cerebellum



Specificity and non-specificity cannot be determined by the acquisition modality but they are function of the radiotracer + knowledge of the clinical question under study

Radiomic features*:

watershed number, variation of contrast, acutance, NIQE, reconstruction matrix dimensions, ...



*Mittal, A., R. Soundararajan, and A. C. Bovik. "Making a Completely Blind Image Quality Analyzer." *IEEE Signal Processing Letters*.

PCA is used for dimensionality reduction. Reduced features have been validated by expert NM visual assessment, are independent on the known clinical variables, show within-center variability, hint at unknown clinical variability and show good harmonization potential.

N. Alchera, *Data harmonization in PET imaging*, Physics PhD dissertation, Univ. of Genova, 2021

towards a better AI

keep an eye on this

safer, trustworthy AI comes from a multidomain effort: contributions from hard and social scientists, legal and economic support, ...



The screenshot shows the European Commission website. At the top, there is the European Commission logo and a search bar. Below the navigation bar, the main heading is "Regulatory framework proposal on artificial intelligence". The page content includes an introductory paragraph, a detailed description of the proposal's goals, and a section titled "Why do we need rules on AI?". On the right side, there is an image of two people looking at a large screen displaying data, and a vertical list of four blue buttons with white text: "Proposal for a Regulation on artificial intelligence", "Impact assessment of the regulation", "Study supporting the impact assessment", and "FAQs: New rules for Artificial".

European Commission

English

Search

Shaping Europe's digital future

Home Policies Activities News Library Funding Calendar Consultations

Home > Policies > Regulatory framework proposal on artificial intelligence

Regulatory framework proposal on artificial intelligence

The Commission is proposing the first-ever legal framework on AI, which addresses the risks of AI and positions Europe to play a leading role globally.

The regulatory proposal aims to provide AI developers, deployers and users with clear requirements and obligations regarding specific uses of AI. At the same time, the proposal seeks to reduce administrative and financial burdens for business, in particular small and medium-sized enterprises (SMEs).

The proposal is part of a wider AI package, which also includes the updated Coordinated Plan on AI. Together, the Regulatory framework and Coordinated Plan will guarantee the safety and fundamental rights of people and businesses when it comes to AI. And, they will strengthen uptake, investment and innovation in AI across the EU.

Why do we need rules on AI?

The proposed AI regulation ensures that Europeans can trust what AI has to offer. While most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that we must address to avoid undesirable outcomes.

For example, it is often not possible to find out why an AI system has made a decision or prediction and taken a particular action. So, it may become difficult to assess whether someone has been unfairly disadvantaged, such as in a hiring decision or in an application for a public benefit scheme.

Although existing legislation provides some protection, it is insufficient to address the specific challenges AI systems may bring.



© gorodenkoff - iStock Getty Images Plus

- Proposal for a Regulation on artificial intelligence >
- Impact assessment of the regulation >
- Study supporting the impact assessment >
- FAQs: New rules for Artificial

Conclusion

- Complex data in medicine require careful preparation before analysis
 - including some significant pre-processing
- Knowledge of the acquisition technique, the clinical case and study conditions are paramount
- Blind, data-driven “muscular” approaches may do wonders on a specific dataset but are rarely trustworthy in a real clinical setting
 - always check on an independent dataset (not just by cross-validation)
 - prefer multi-centric studies to improve robustness
- My experience favours knowledge-driven features/algorithm
 - embed the clinical, technical and physiological information
 - easily test significance and relate to literature
 - natural model explainability

