Artificial intelligence in personalized medicine

Enrico Glaab
Luxembourg Centre for Systems Biomedicine

Overview

(1) Introduction: Artificial Intelligence vs. Classical Statistics

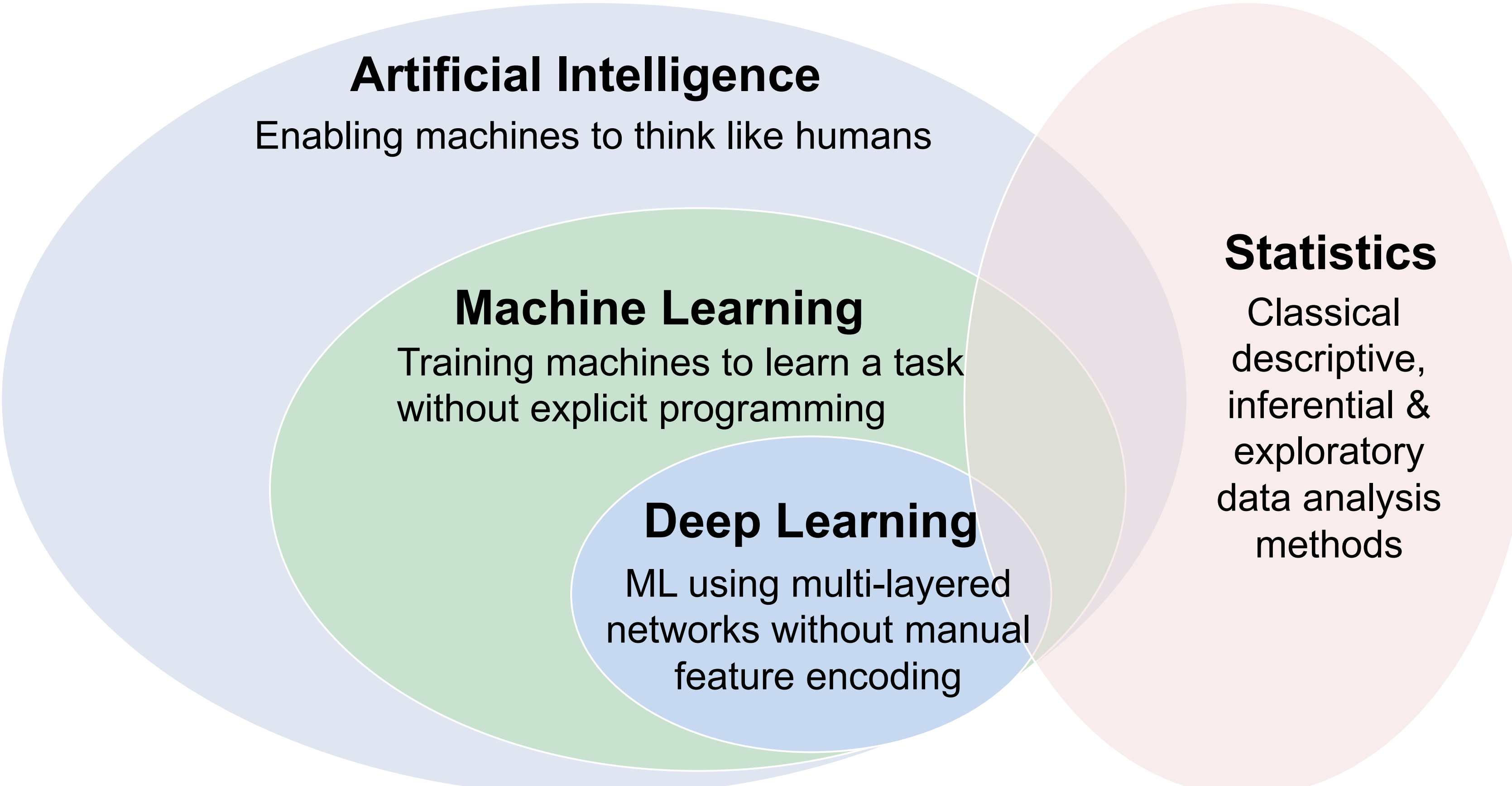
(2) AI success stories in recent years

(3) AI in personalized medicine:
- Biomarker discovery
- Drug discovery
- Digital health monitoring

(4) Common pitfalls and challenges

(5) Outlook on new AI strategies to address the challenges

AI vs. Statistics

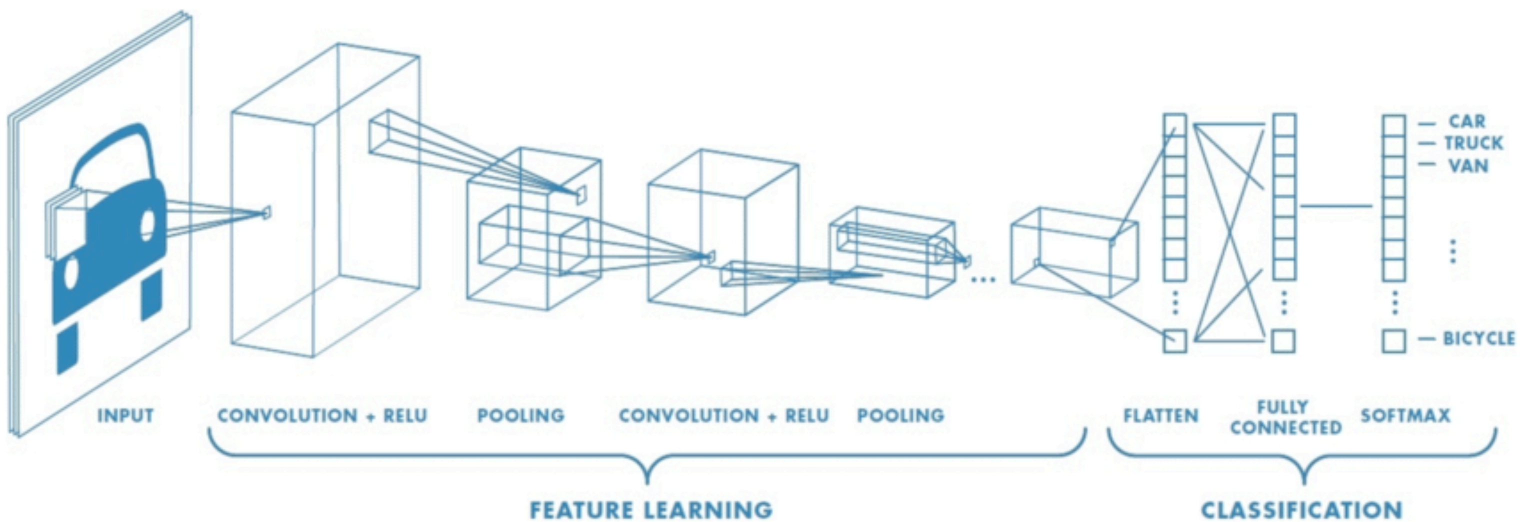


UNIVERSITÉ DU LUXEMBOURG

LCSB

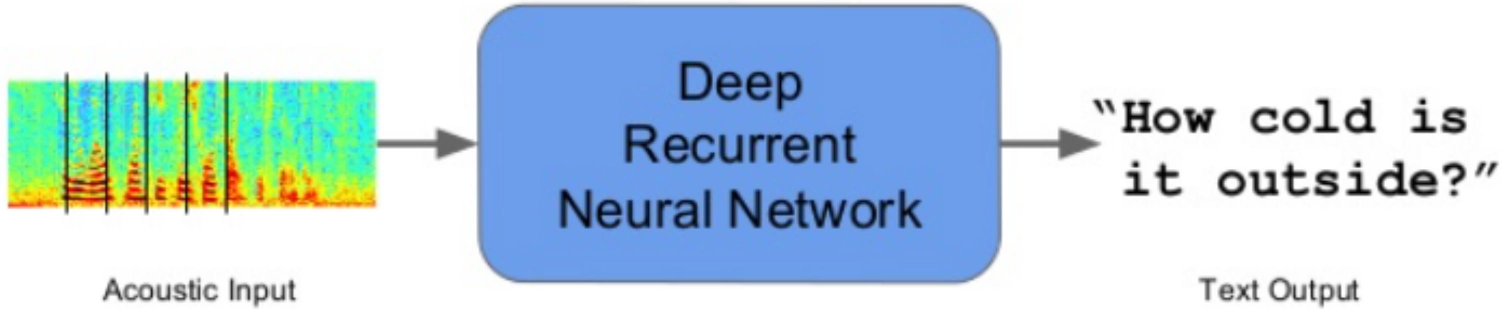
AI success stories: Image data (1)

- **Image classification**: unprecedented accuracies using deep learning (e.g. “AlexNet” approach by Krizevsky et al., 2012)
- **Novel techniques**: Convolutional Nets / Transformers, Dropout, Data Augmentation, Rectifier Nonlinearity, Local Response Normalization



AI success stories: Sequential data (2)

- **Speech recognition**: 30% less errors with new neural networks (Huang et al., ACM Communications, 2012)
- **Novel techniques**: Recurrent Neural Networks (RNNs): Long-Short-Term Memory Networks (LSTMs), Gated Recurrent Units (GRU), etc.



AI successes in personalized medicine - Biomarkers

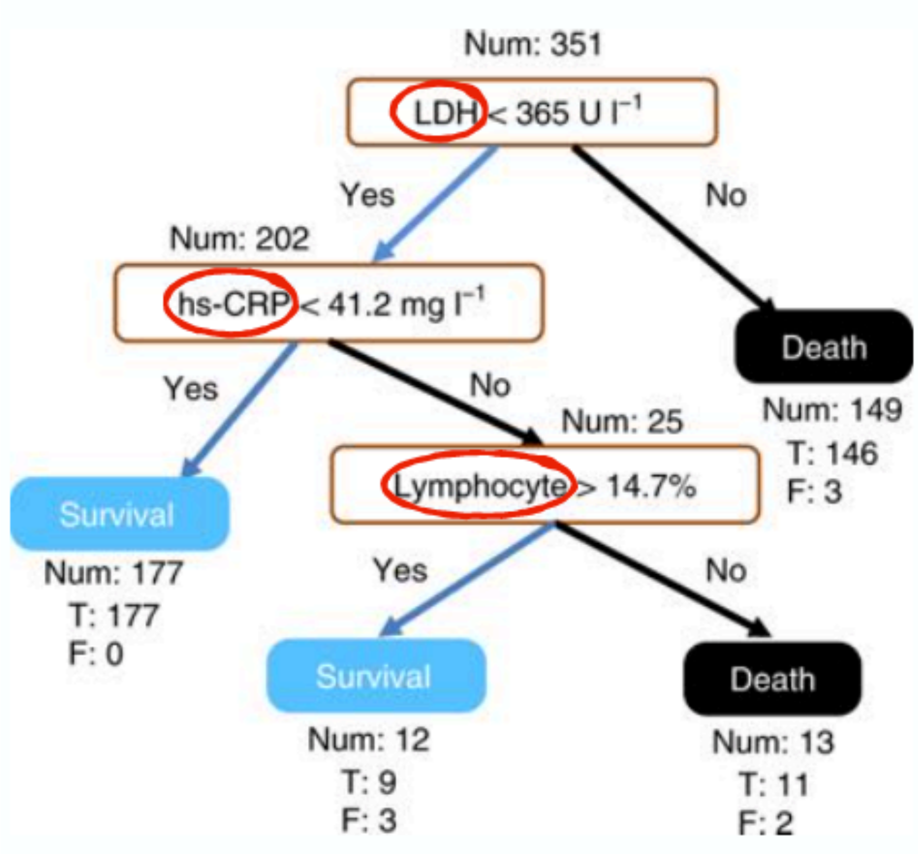
Multiple omics-based AI diagnostic tools already clinically validated:

Name	Test approval (FDA-cleared and/or LDT)	Purpose	References
MammaPrint	FDA-cleared, LDT	breast cancer risk-of-recurrence assessment	Van't Veer et al., Nature, 2002
AlloMap Heart	FDA-cleared, LDT	identifying heart transplant recipients with risk of cellular rejection	Yamani et al., J Heart Lung Transplant, 2007
Prosigna Assay / PAM50	FDA-cleared, LDT	breast cancer risk of distant recurrence prediction	Nielsen et al., BMC Cancer, 2014
Oncotype DX	LDT	breast cancer risk-of-recurrence assessment	Kelley et al., Cancer, 2010
Decipher	LDT	prostate cancer metastatic risk prediction	Marrone et al., PLoS Curr., 2015

AI in biomedicine: Covid-19 hospital mortality prediction

- Covid-19: alterations in common clinical blood tests when diagnosed

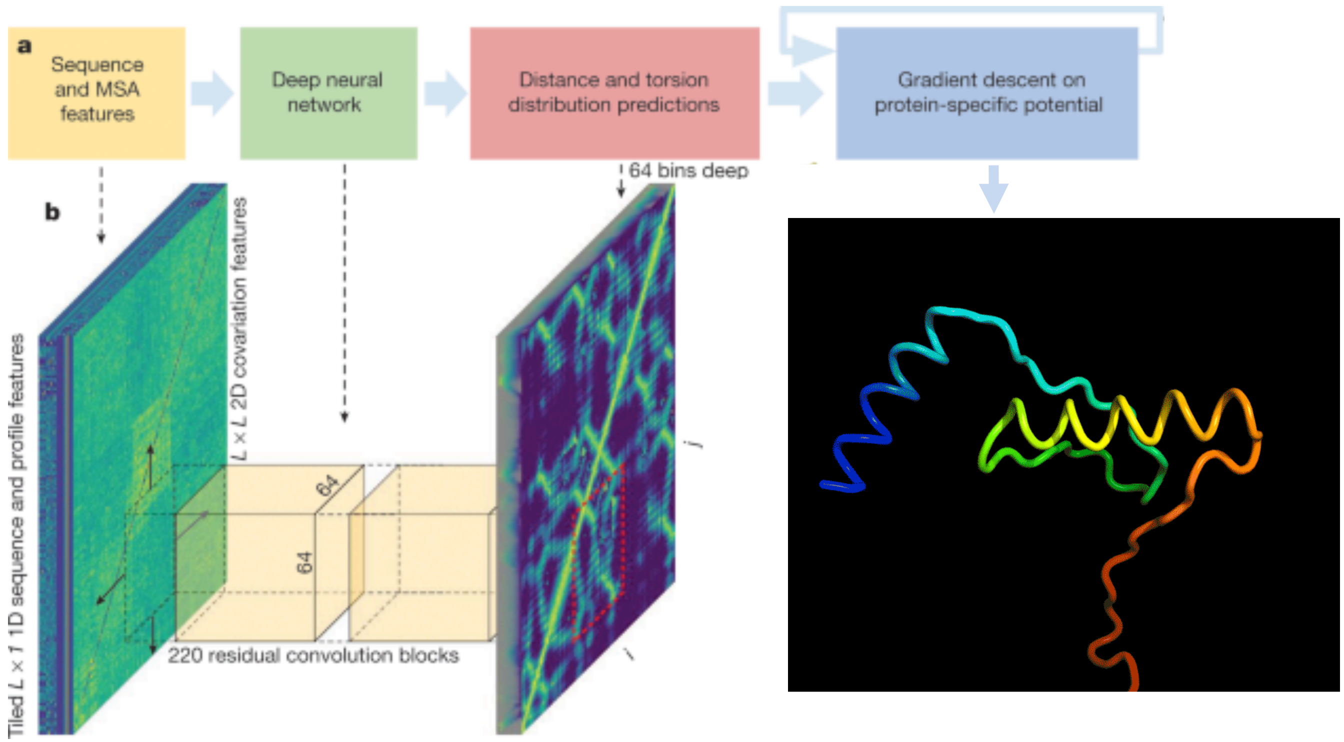
- Apply ML → 90% accurate predictions of mortality on test set of 110 patients (10 days in advance)



(source: Yan et al., Nat. Mach. Intell. 2, 2020)

Deep learning for biomedical research: AlphaFold 2

- DeepMind's AlphaFold 2 → major advance in protein structure prediction
- Scores > 90 in global distance test (GDT) in the CASP competition



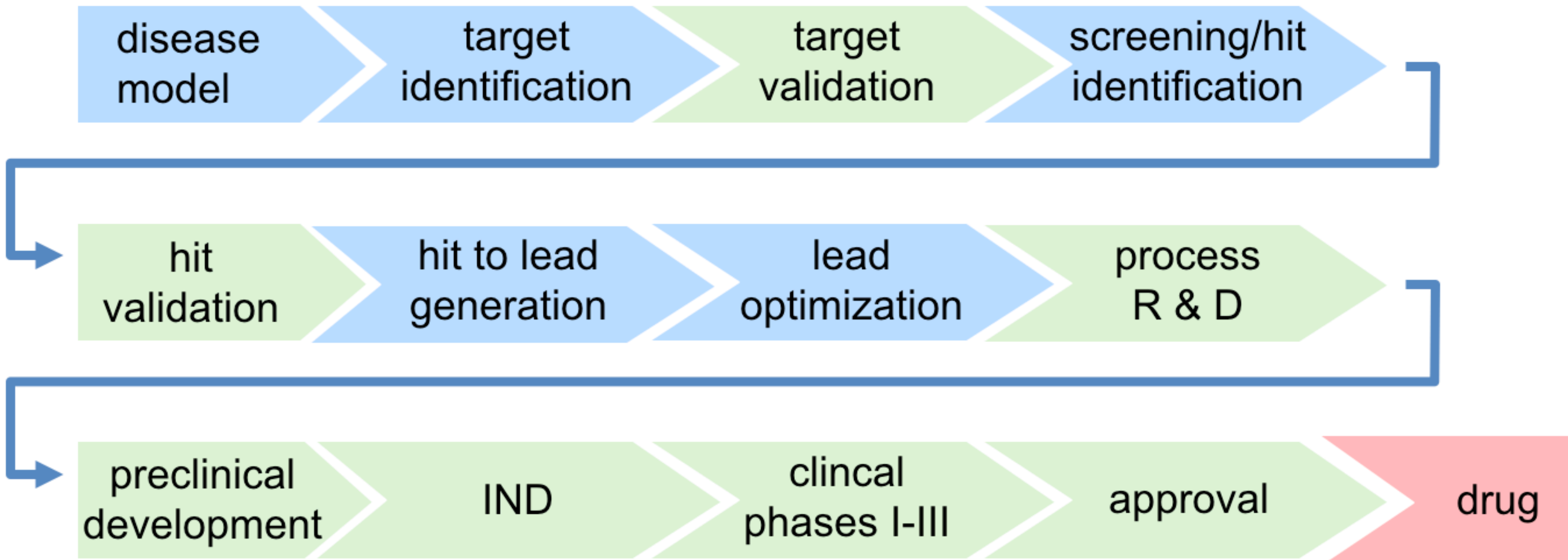
(AlphaFold basic architecture; source: Senior et al., Nature 577, 2020)

UNIVERSITÉ DU LUXEMBOURG

LCSB

AI in drug development

Common phases in drug development:



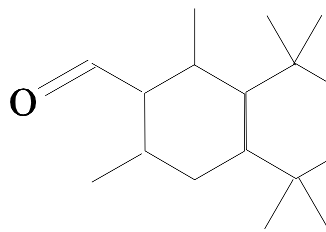
Opportunities for AI approaches highlighted in blue ( )

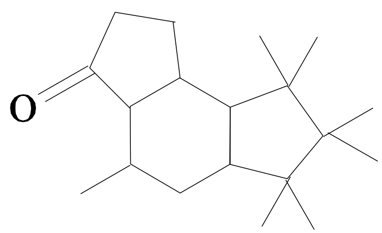
AI in drug development

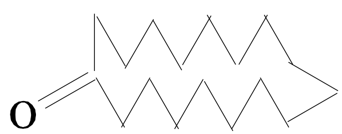
- **Question**: Which drug-like compounds bind to a particular target protein?

- **Example**: Active compounds Inactive / low activity compounds

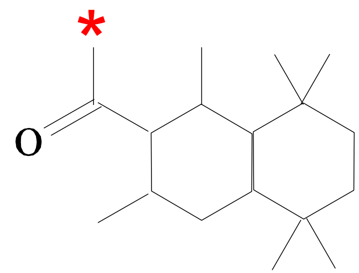
Active compounds

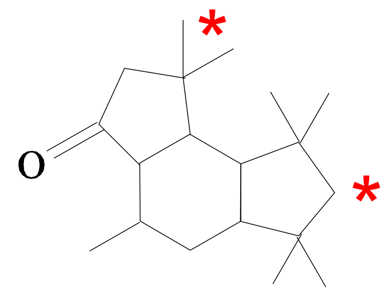






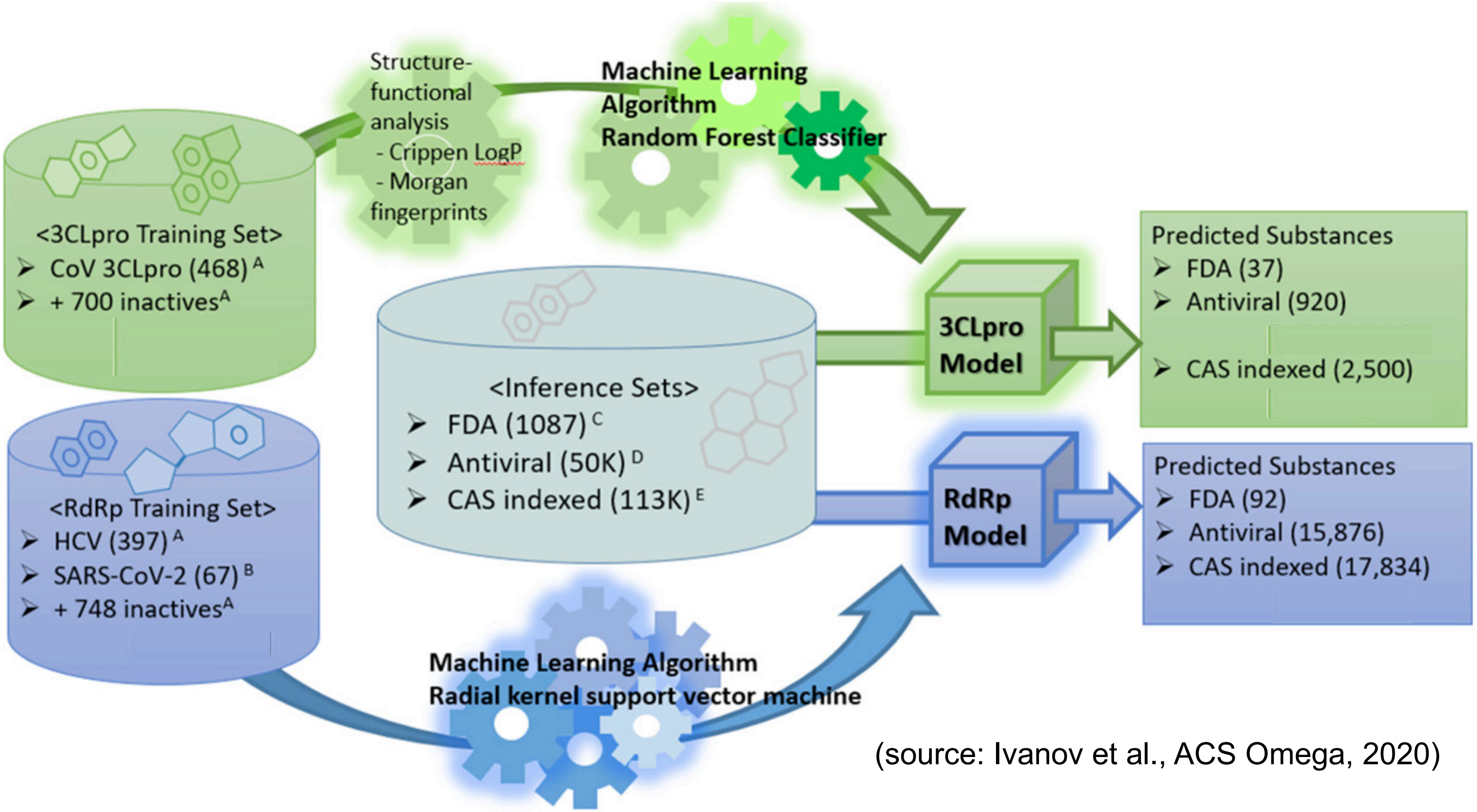
Inactive / low activity compounds





Application to SARS-CoV-2 compound screening

AI-based screening: Finding new inhibitors of viral proteins *3CLpro* and *RdRp*



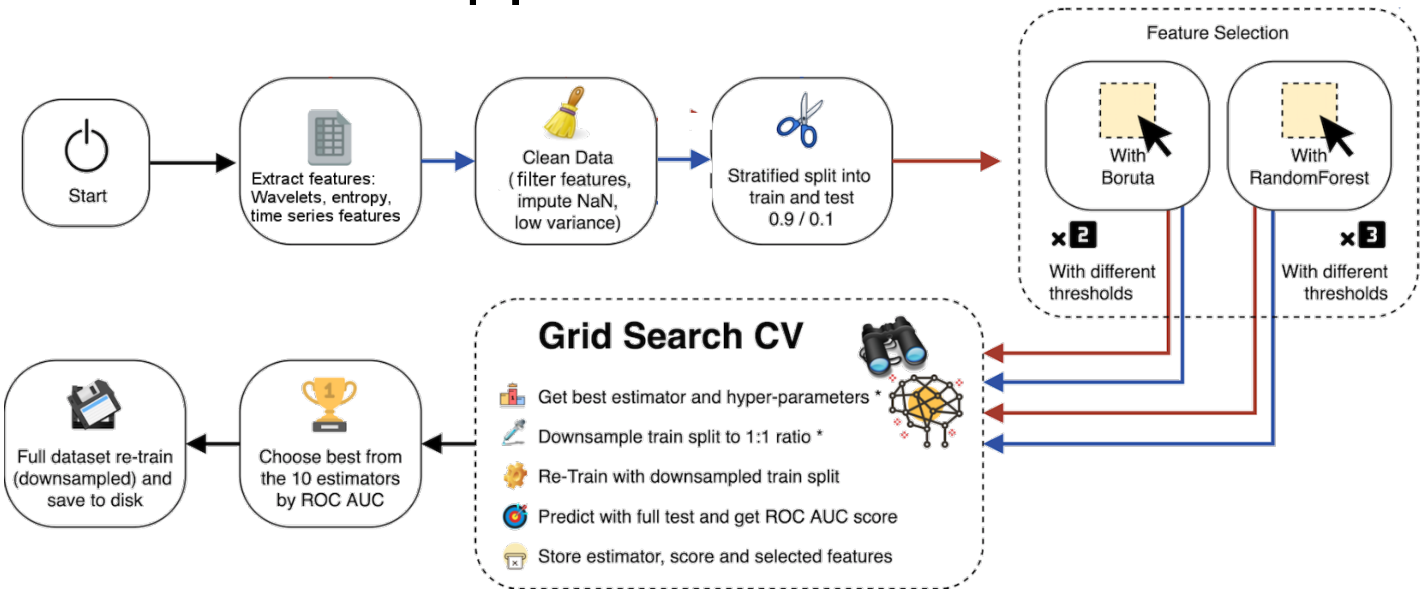
(source: Ivanov et al., ACS Omega, 2020)

AI for personalized health monitoring / digital biomarkers

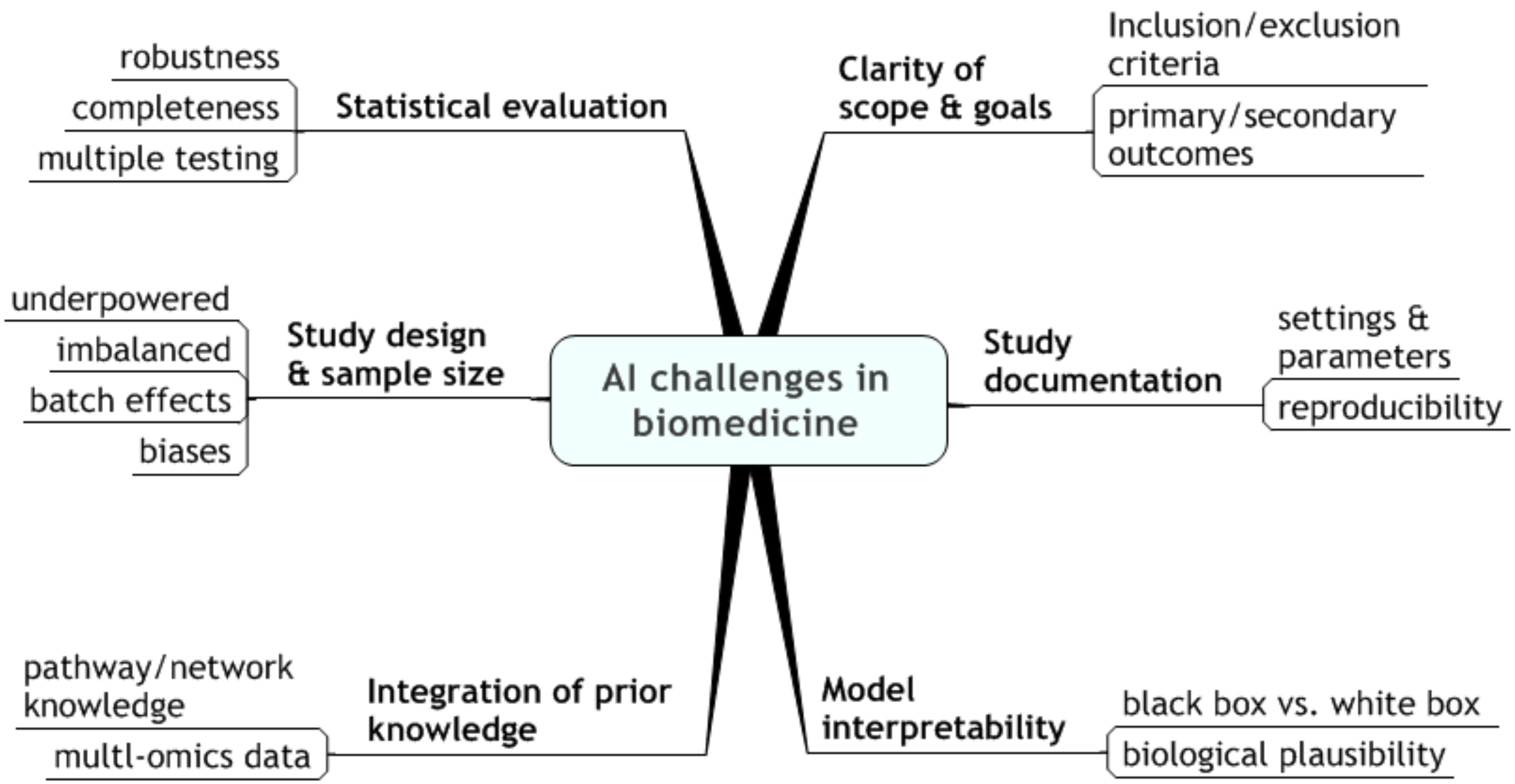
Digital biomarkers:

- mobile phone gyrometer & accelerometer
- gait sensors (eGaIT system)

Feature extraction & ML pipeline:



Common challenges for AI in personalized medicine



Recommendations from literature

Data pre-processing, filtering & normalization:

→ use cross-validation to check if pre-processing leads to information loss

→ compare or combine multiple pre-processing approaches

Integration of prior knowledge & multi-omics analyses:

→ Assess cost/benefit of multi-omics analyses using prior data, or conduct pilot analyses

→ use existing software & frameworks for integrative biological data analysis

Ensuring model interpretability & biological plausibility:

→ use dedicated methods to build interpretable models (e.g. rule learning methods)

→ use cellular pathway/network analysis & literature mining to guide modeling

Outlook: New AI strategies to address the challenges

Increase statistical power

→ AI-based integration of prior knowledge & multi-omics data

→ Algorithmic sample matching & selection, AI for batch adjustment

Increase model robustness

→ AI-based integration of models across human & animal model data, different experimental platforms and cohorts (Transfer learning)

Increase model interpretability

→ Structured and graph-based machine learning

→ AI-based literature mining for model interpretation

Acknowledgements

Biomedical Data Science Group

New members:

Mirko Ledda

Elisa Gomez de Lope

Loïc Le Bescond

Fonds National de la Recherche Luxembourg

FONDATION DE LUXEMBOURG Philanthropy in action

Fondation Gustave et Simone Prévot en faveur de la Recherche Médicale

ERACoSysMed

ERAPerMed

ACTIVE MOTIF

European Commission HORIZON 2020

UNIVERSITÉ DU LUXEMBOURG

LCSB

Thank you for your attention!

Questions?

Comments?

References

1. E. Glaab, *Building a virtual ligand screening pipeline using free software: a survey*, Briefings in Bioinformatics (2015), 17(2), pp. 352
2. E. Glaab, *Using prior knowledge from cellular pathways and molecular networks for diagnostic specimen classification*, Briefings in Bioinformatics (2015), 17(3), 440
3. N. Vlassis, E. Glaab, *GenePEN: analysis of network activity alterations in complex diseases via the pairwise elastic net*, Statistical Applications in Genetics and Molecular Biology (2015), 14(2), 221
4. E. Glaab, J. M. Garibaldi, N. Krasnogor. *Learning pathway-based decision rules to classify microarray cancer samples*, German Conference on Bioinformatics 2010, Lecture Notes in Informatics (LNI), 173, 123-134
5. E. Glaab, J. Bacardit, J. M. Garibaldi, N. Krasnogor, *Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data*, PLoS ONE, 7(7):e39932, 2012
6. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. Extending pathways and processes using molecular interaction networks to analyse cancer genome data, BMC Bioinformatics, 11(1):597, 2010
7. E. Glaab, A. Baudot, N. Krasnogor, R. Schneider, A. Valencia. EnrichNet: network-based gene set enrichment analysis, Bioinformatics, 28(18):i451-i457, 2012
8. Maes, M., Nowak, G., Caso, J. R., Leza, J. C., Song, C., Kubera, M., .et al. (2016). Toward omics-based, systems biomedicine, and path and drug discovery methodologies for depression-inflammation research. Molecular neurobiology, 53(5), 2927-2935.
9. E. Glaab, J.P. Trezzi, A. Greuel, C. Jäger, Z. Hodak, A. Drzezga, L. Timmermann, M. Tittgemeyer, N. J. Diederich, C. Eggers, Integrative analysis of blood metabolomics and PET brain neuroimaging data for Parkinson's disease, Neurobiology of Disease (2019), Vol. 124, No. 1, pp. 555
10. E. Glaab, R. Schneider, *Comparative pathway and network analysis of brain transcriptome changes during adult aging and in Parkinson's disease*, Neurobiology of Disease (2015), 74, 1-13
11. Z. Zhang, P. P. Jung, V. Grouès, P. May, C. Linster, E. Glaab, *Web-based QTL linkage analysis and bulk segregant analysis of yeast sequencing data*, GigaScience (2019), 8(6), 1-18
12. S. Köglsberger, M. L. Cordero-Maldonado, P. Antony, J. I. Forster, P. Garcia, M. Buttini, A. Crawford, E. Glaab, *Gender-specific expression of ubiquitin-specific peptidase 9 modulates tau expression and phosphorylation: possible implications for tauopathies*, Molecular Neurobiology (2017), 54(10), pp. 7979
13. Kleiderman, S., Gutbier, S., Ugur Tufekci, K., Ortega, F., Sá, J. V., Teixeira, A. P., et al. (2016). Conversion of Nonproliferating Astrocytes into Neurogenic Neural Stem Cells: Control by FGF2 and Interferon-γ. Stem Cells, 34(12), 2861-2874.
14. Bolognin, S., Fossépré, M., Qing, X., Jarazo, J., Ščančar, J., Moreno, E. L., et al. (2019). 3D Cultures of Parkinson's Disease-Specific Dopaminergic Neurons for High Content Phenotyping and Drug Testing. Advanced Science, 6(1), 1800927.
15. Jaeger, C., Glaab, E., Michelucci, A., Binz, T. M., Koeglsberger, S., Garcia, P., ... & Buttini, M. (2015). The mouse brain metabolome: region-specific signatures and response to excitotoxic neuronal injury. The American Journal of Pathology, 185(6), 1699-1712.
16. E. Glaab, R. Schneider, *RepExplore: Addressing technical replicate variance in proteomics and metabolomics data analysis*, Bioinformatics (2015), 31(13), pp. 2235
17. E. Glaab, R. Schneider, *PathVar: analysis of gene and protein expression variance in cellular pathways using microarray data*, Bioinformatics, 28(3):446-447, 2012
18. E. Glaab, A. Baudot, N. Krasnogor, A. Valencia. *TopoGSA: network topological gene set analysis*, Bioinformatics, 26(9):1271-1272, 2010
19. E. Glaab, J. M. Garibaldi and N. Krasnogor. *ArrayMining: a modular web-application for microarray analysis combining ensemble and consensus methods with cross-study normalization*, BMC Bioinformatics,10:358, 2009