Introduction to Real-World Data and Artificial Intelligence in Healthcare and Drug Development

Yahui Tian, Ph.D.
Principal Therapeutic Area and Methodology Statistician
Boehringer Ingelheim, Ridgefield
ASA CT Chapter Mini Conference, April 14, 2023

Boehringer Ingelheim

Outline

- RWD
  - Concepts
  - Regulatory initiatives
  - Applications in drug development
- Artificial Intelligence
  - Machine learning
    - ✓ The logic&models
    - ✓ Applications in drug development and healthcare
    - ✓ One use case
  - Deep learning
    - ✓ Network structure
    - ✓ Advantages in analysing unstructured data
- Summary&Discussion

RWD Concepts

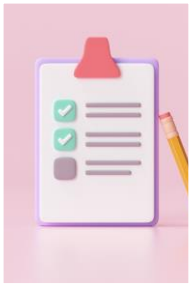
FDA Guidance: Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drug and Biological Products Guidance for Industry

Real-world data

data relating to patient health status and /or the delivery of health care that are routinely collected from a variety of sources.

Real-world evidence

clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

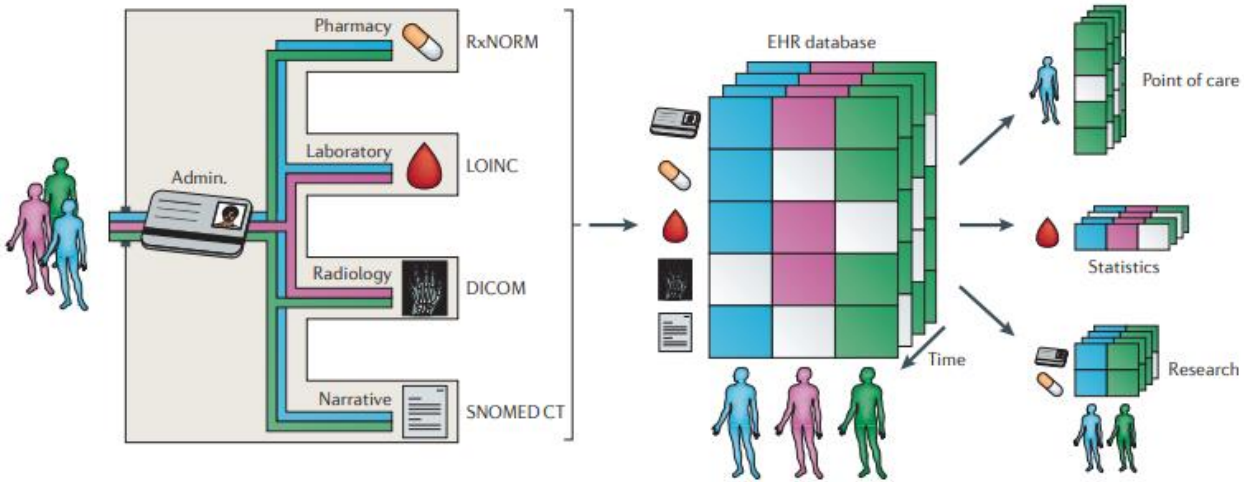


Sources of real-world data

- Electronic health record (EHR) data
- Medical claims data
-Product or disease registry data
- Data obtained from digital health technologies
- Data gathered from other sources that can inform on health status, such as questionnaires

EHR/Claims Databases

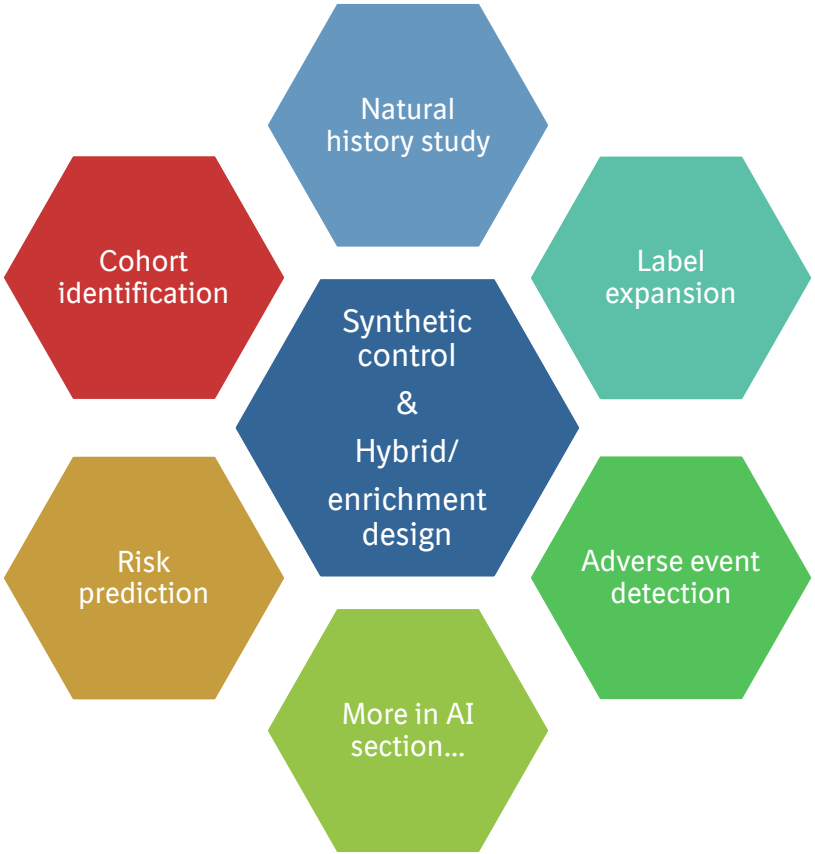
- Primary goal: document patients' care, reimbursement
- Data format
- Advantages: large sample size, long follow-up, cost-effective and time-saving source of research
- Obstacles: censoring, irregular time series data, completeness, correctness and confounding effects



Note: picture comes from: Jensen, Peter B., Lars J. Jensen, and Søren Brunak. "Mining electronic health records: towards better research applications and clinical care." *Nature Reviews Genetics* 13.6 (2012): 395-405.

Regulatory-supported initiatives (selected)	
Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products	Published recently for comment purpose. Provides recommendations to sponsors and investigators considering the use of externally controlled clinical trials to provide evidence of the safety and effectiveness of a drug product.
Guidance to submit documents using real-world data and real world evidence to FDA for drugs and biologics	To encourage sponsors and applicants to generate RWE as part of a regulatory submission to FDA
Framework for FDA: Real-world evidence program	To better evaluate the potential use of RWE in supporting new indications for drugs already approved or for post-approval study. Demonstrate a commitment to transforming drug development cycle.
Guidance on the use of Electronic Health Records (FDA)	Focus on data integrity.
RCT DUPLICATE (FDA funded)	Brigham and Women's hospital/Harvard Medical School/Aetion. Try to replicate results from 30 RCTs using RWD. Similar study: REPEAT, REPEAT - Home (repeatinitiative.org)
HMA/EMA joint task force on big data	To understand the acceptability of RWE from "Big data" to support regulatory evaluation and monitoring
Clinical trials transformation initiative (FDA)	FDA and Duke University. Encourage the increased use of RWD. Sentinel IMPACT-Afib (Atrial Fibrillation), cluster-randomized pragmatic study.
Regulatory update for promoting RWD utilization (PMDA)	The next 5 years will be an important period for considering the expanding the use of RWD to support regulatory process.
Key Considerations in Using Real-World Evidence to Support Drug Development (NMPA)	How to use RWD and RWE as complementary evidence to RCT in evaluating efficacy and safety
Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan (FDA)	Outline regulate machine learning algorithm under the software as a medical device framework.

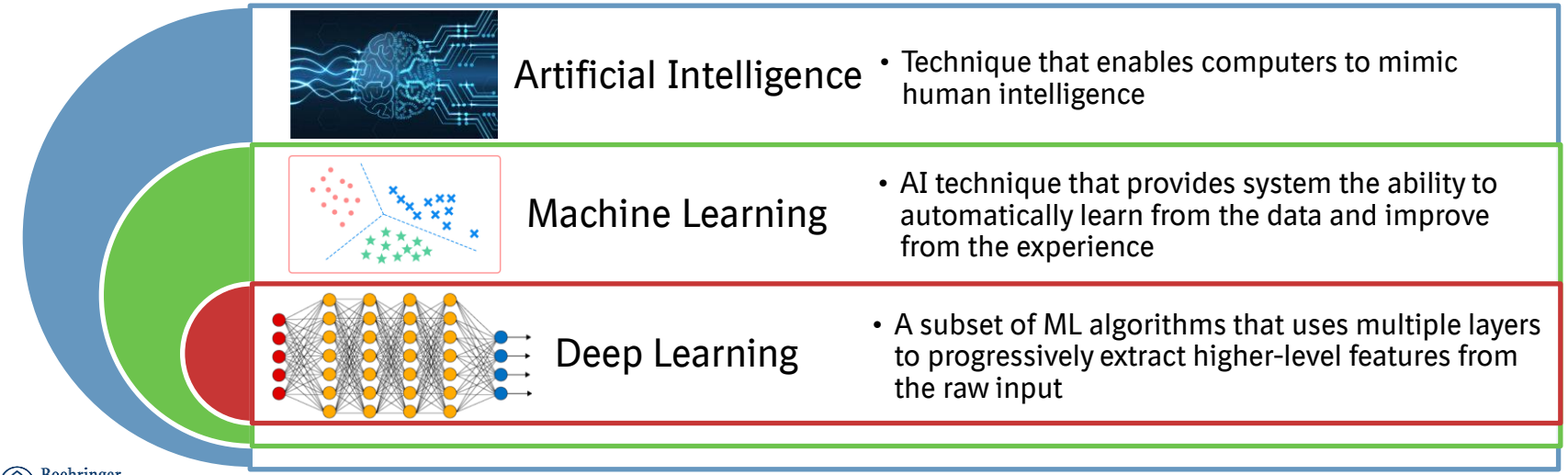
Applications of RWD in clinical development



AI Family

Importance of artificial intelligence in medicine and healthcare

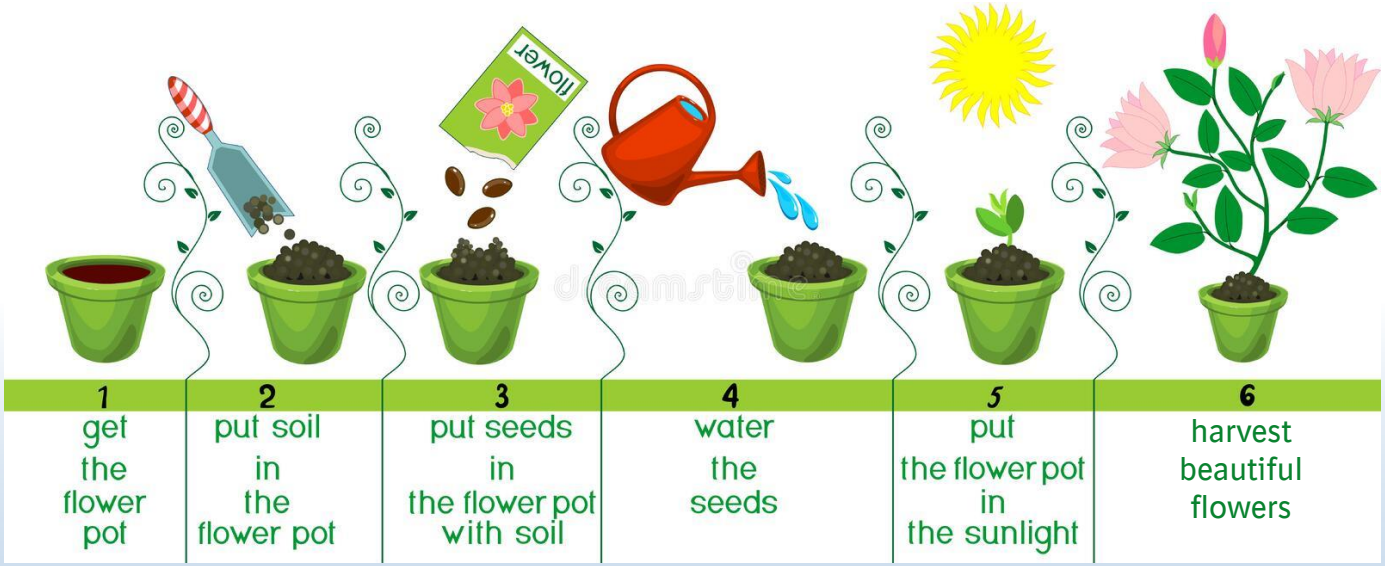
- Dive into “deep” data and discover unknown: features discovered through data driven process
- Dramatically increase of AI related regulatory submissions from 2016 to 2021
- AI have achieved great success in many domains including disease progression, disease screening, treatment response prediction, risk stratification and etc..



The Logic

Mimic dynamics in the Nature

Gardening



1	2	3	4	5	6
get the flower pot	put soil in the flower pot	put seeds in the flower pot with soil	water the seeds	put the flower pot in the sunlight	harvest beautiful flowers

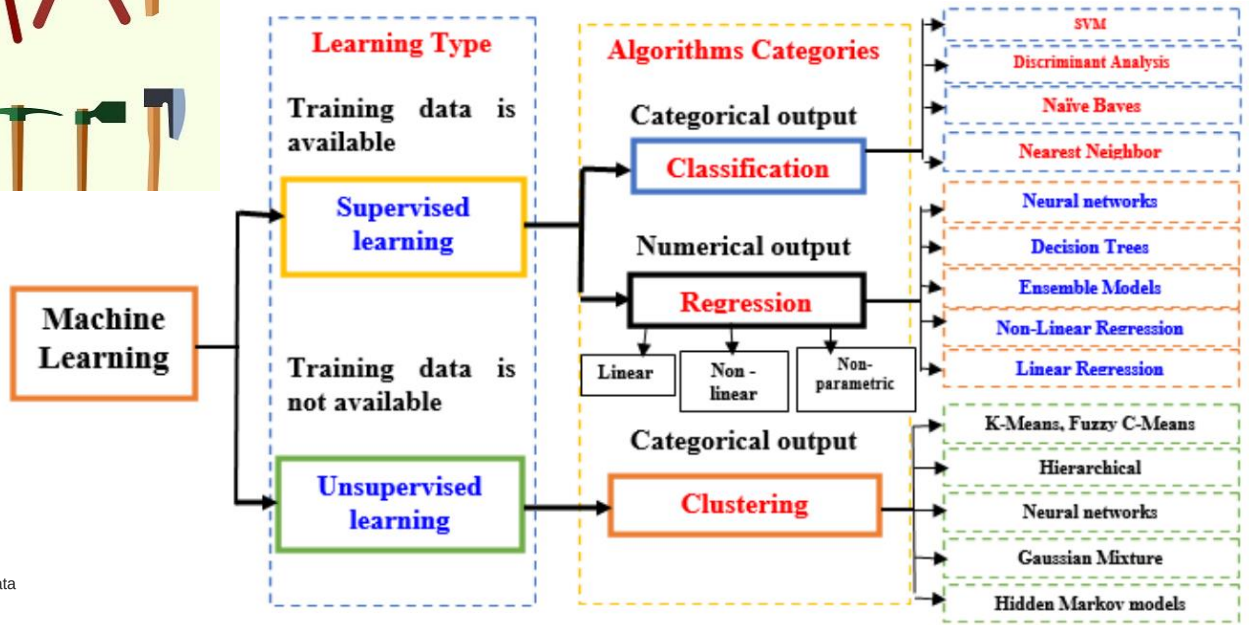
Machine Learning

1	2	3	4 & 5	6
build data matrix	put cleaned data in the matrix	build initial model/ algorithm	nurturing: • model tunning • hyperparameter optimization • error analysis…	get insightful outputs

Toolbox

Tools for gardening

Machine Learning Toolbox



ML picture from: Kumar, Dileep, et al. "Social Sentimental Analytics using Big Data Tools." *Communication and Power Engineering* (2017): 266.

What ML can bring to us?

Weissler *et al. Trials* (2021) 22:537
https://doi.org/10.1186/s13063-021-05489-x

Trials

COMMENTARY Open Access

The role of machine learning in clinical research: transforming the future of evidence generation

E. Hope Weissler[1*], Tristan Naumann[2], Tomas Andersson[3], Rajesh Ranganath[4], Olivier Elemento[5], Yuan Luo[6], Daniel F. Freita
Emmette Hutc
Erich Huang[1,1]

PERSPECTIVES

PERSPECTIVE

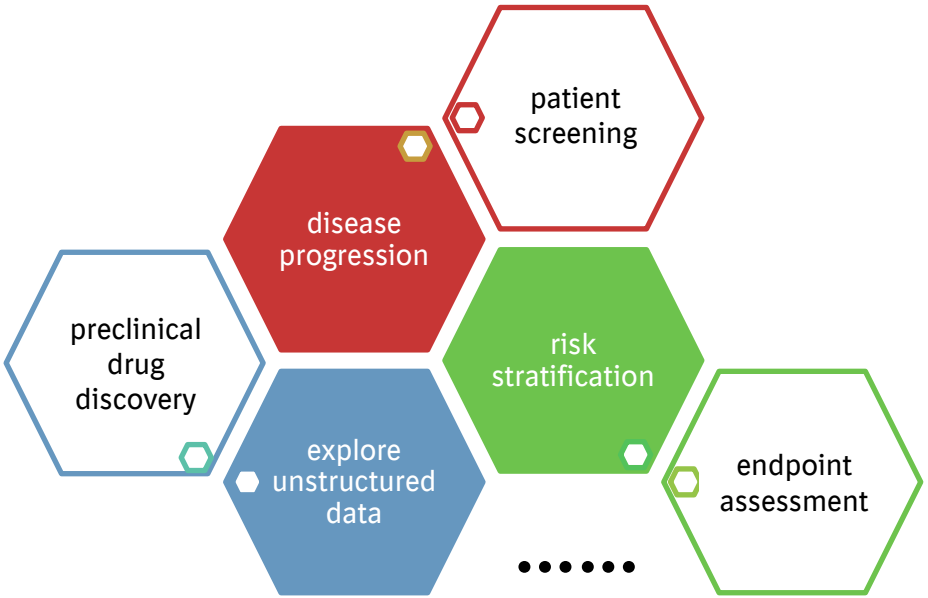
Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021

Qi Liu[1,†], Ruihao Huang[1,†], Julie Hsieh[1,†], Hao Zhu[1,*,†], Mo Tiwari[1], Guansheng Liu[1], Daphney Jean[1], M. Khair ElZarrad[2], Tala Fakhouri[2], Steven Berman[3], Billy Dunn[3], Matthew C. Diamond[4] and Shiew-Mei Huang[1]

envisioned that AI/ML would play an increasingly important role in drug development.[1] That prediction has now been confirmed by this landscape analysis based on drug and biologic regulatory submissions to the FDA from 2016 to 2021.

THE TREND OF INCREASING AI/ML–RELATED SUBMISSIONS AT THE FDA'S CENTER FOR DRUG EVALUATION AND RESEARCH

This analysis was performed by searching for submissions with key terms "machine learning" or "artificial intelligence" in Center for Drug Evaluation and Research (CDER) internal databases for Investigational New Drug applications, New Drug Applications, Abbreviated New Drug Applications, and Biologic License Applications, as well as submissions for Critical Path Innovation



ML use case

- **Goal:** To seek emergency use authorization (EUA) of anakinra for the treatment of COVID-19 in hospitalized patients with pneumonia requiring supplemental oxygen (low or high-flow oxygen) who are at risk of progressing to severe respiratory failure.
- **Why AI/ML:** Patients in randomized, double-blinded trials were selected by a biomarker test (suPAR), but not commercially available in the United States. Need to explore surrogate characteristics to reliably identify target population.
- **Algorithm design:**
  - Outcome: suPAR>=6 ng/mL, yes/no
  - Candidate factors
  - Methods: elastic net regression and neural network model
  - Evaluation: PPV is more important

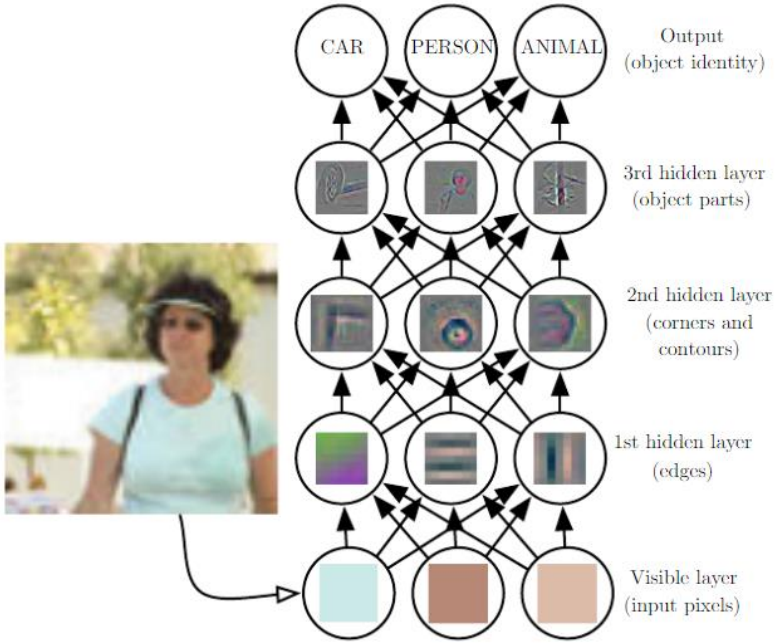
Emergency Use Authorization (EUA) for Kineret (Anakinra)
THE UNAPPROVED USE OF AN APPROVED PRODUCT
Center for Drug Evaluation and Research (CDER) Review

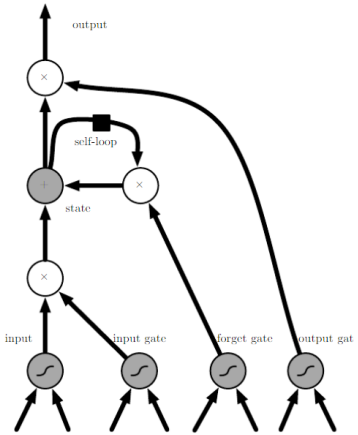
Identifying Information

Application Type (EUA or Pre-EUA) If EUA, designate whether pre-event or intra-event EUA request.	EUA
EUA Application Number(s)	109
Sponsor (entity requesting EUA or pre-EUA consideration), point of contact, address, phone number, fax number, email address	Swedish Orphan Biovitrum AB (Sobi) c/o Advyzom LLC. 335 Snyder Ave. Berkeley Heights, NJ 07922 Rula Ibrahim-Saker, PharmD US Agent

- **Value&Benefits:**
  - Identified criteria with high PPV
  - Results are consistent with medical understanding of inflammation and immune status of patients
  - New criteria are associated with higher risk of developing severe respiratory failure

“Deep” Learning





Forget gate $f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right),$

State $s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right),$

Input gate $g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right)$

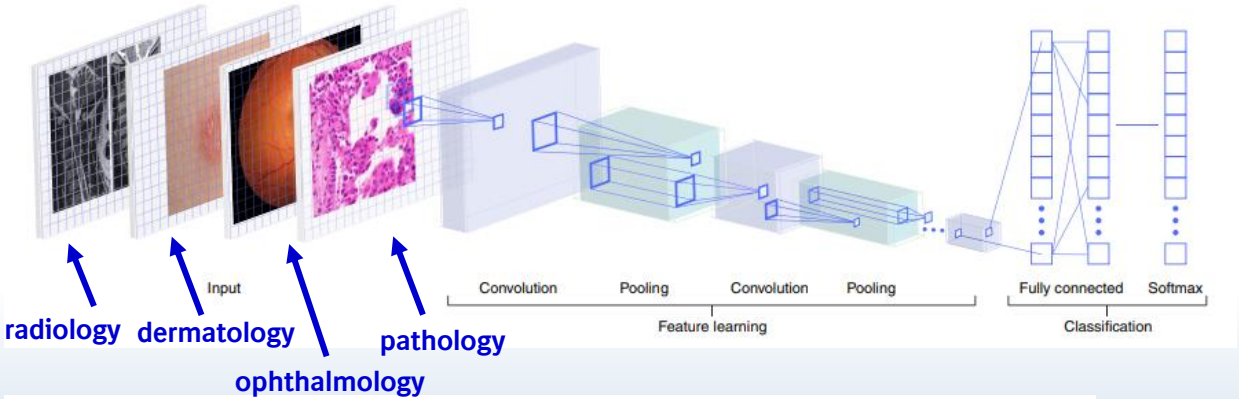
Output $h_i^{(t)} = \tanh\left(s_i^{(t)}\right) q_i^{(t)},$

Output gate $q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right),$

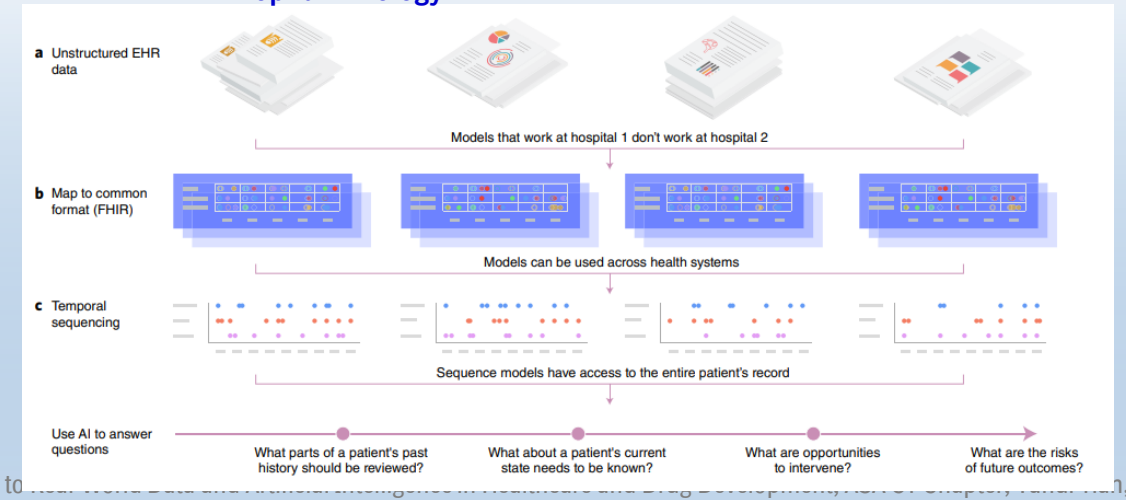
Pictures from: Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

Advances of DL in understanding and manipulating unstructured data

Medical Imaging



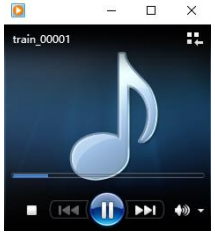
Medical Notes

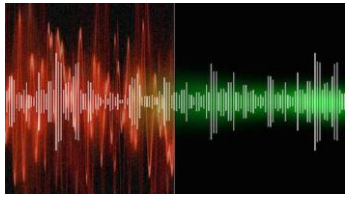


Pictures from: Esteva, Andre, et al. "A guide to deep learning in healthcare." *Nature medicine* 25.1 (2019): 24-29.

Audio Data Analysis

Listen to the audio

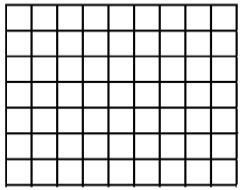




Pros:
- Easy to start
- Good interpretability

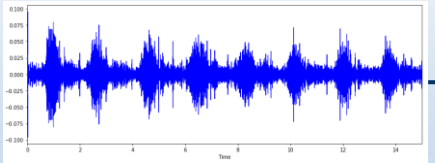
Features

Samples

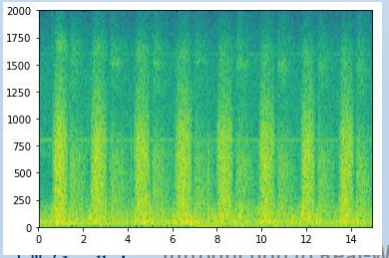


Look at data

Time domain



Frequency domain



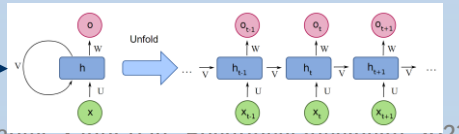
Filters & Pre-processing

Select algorithm

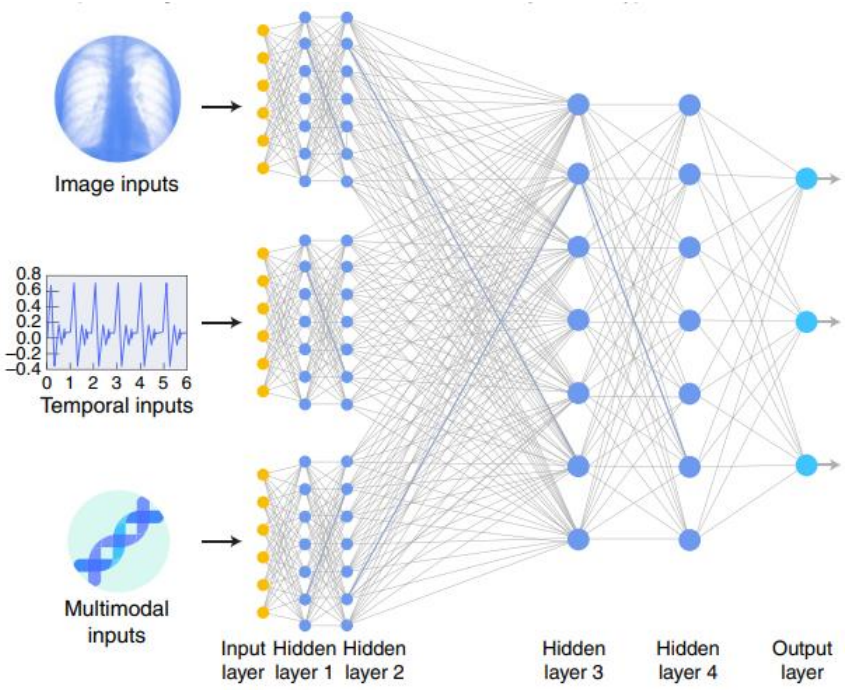
Traditional Approach: feature engineering with domain knowledge + machine learning algorithms

Deep learning:
Data augmentation + transfer learning
- Sound→image→CNN
- Sound→RNN

Pros:
- More advanced algorithm
- More computation power
- Transfer learning



Recent Progress



Pictures from: Esteva, Andre, et al. "A guide to deep learning in healthcare." *Nature medicine* 25.1 (2019): 24-29.

ARTICLES
https://doi.org/10.1038/s42256-021-00357-4
nature machine intelligence
Check for updates

Deep learning prediction of patient response time course from early data via neural-pharmacokinetic/ pharmacodynamic modelling

James Lu[1], Brendan Bender[1], Jin Y. Jin[1] and Yuanfang Guan[2]

nature medicine
REVIEW ARTICLE
https://doi.org/10.1038/s41591-022-01981-2
Check for updates

Multimodal biomedical AI

Julián N. Acosta[1], Guido J. Falcone[1], Pranav Rajpurkar[2,4] and Eric J. Topol[3,4]

Harnessing multimodal data integration to advance precision oncology

Kevin M. Boehm, Pegah Khosravi, Rami Vanguri, Jianjiong Gao and Sohrab P. Shah

Summary&Discussion

- RWD and AI to unlock data power in drug development and healthcare
- Define question clearly
- Data limitation&standardization
- Explainable AI
- Multidisciplinary collaboration

THANK YOU

yahui.tian@boehringer-ingelheim.com

Boehringer Ingelheim