P1 Guest Talk

All that glitters is not gold: Tales from applying deep learning to clinical data

Chloé-Agathe Azencott

Centre for Computational Biology (CBIO)
Mines Paris – PSL, Institut Curie & INSERM U1331
PR[AI]RIE-PSAI

May 21, 2025

http://cazencott.info chloe-agathe.azencott@minesparis.psl.eu @cazencott@lipn.info

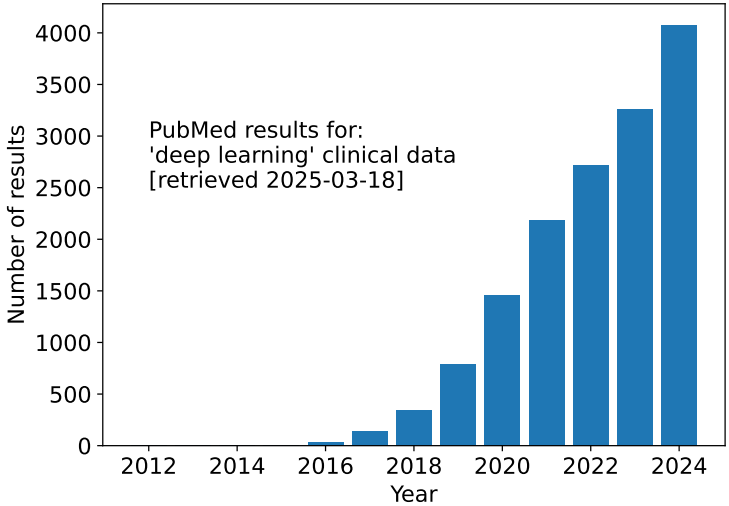
MINES PARIS | PSL

PR[AI]RIE Paris School

institut Curie

Inserm

Deep learning is everywhere



Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens

Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens

Sometimes

Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens
- the **nature of the data** is well understood

  $\Rightarrow$ good modeling/architecture

Sometimes

Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens

Sometimes

- the **nature of the data** is well understood

  $\Rightarrow$ good modeling/architecture

Tabular vs unstructured data

Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens

Sometimes

- the **nature of the data** is well understood

  $\Rightarrow$ good modeling/architecture

Tabular vs unstructured data

- the **nature of the problem** is well understood

  humans can do it

Deep learning works particularly well when

- the data is **really big**
  - ImageNet: 14 million images
  - LLama [Tou+23]: 2 trillion tokens

Sometimes

- the **nature of the data** is well understood

  $\Rightarrow$ good modeling/architecture

Tabular vs unstructured data

- the **nature of the problem** is well understood

  humans can do it

Sometimes

Deep learning does not work so well on tabular data

- **Tabular/structured** data: organized as a table of values

  omics data, lab measurements, vital signs, clinical data, etc.

Deep learning does not work so well on tabular data

- **Tabular/structured** data: organized as a table of values

  omics data, lab measurements, vital signs, clinical data, etc.

- **Tree-based** models remain state-of-the-art on tabular data [GOV22; McE+23]
  - Robustness to **uninformative features** and **irregular features** (skewed, high variance, etc)
  - Ability to learn **non-smooth** functions
  - Non-invariance by **rotation** [Ng04]

Deep learning does not work so well on tabular data

- **Tabular/structured** data: organized as a table of values
  omics data, lab measurements, vital signs, clinical data, etc.

- **Tree-based** models remain state-of-the-art on tabular data [GOV22; McE+23]
  - Robustness to **uninformative features** and **irregular features** (skewed, high variance, etc)
  - Ability to learn **non-smooth** functions
  - Non-invariance by **rotation** [Ng04]

- A possible solution: **pretraining**
  - learn good **representations** on very large data sets; **fine-tune** on the task at hand
  - **CARTE** [KGV24] and **TabPFN** [Hol+25]

Deep learning is better suited to (big) non-tabular data

- **Images**

  [Gul+16] learns from 128 175 retinal fundus photographs

- **Graphs**

  [LU21] predicts chronic disease by creating a patient network from 1.2 million EHRs

- **Longitudinal data**
  - [Raj+18] learns medical events (unplanned readmission, mortality) from 216 221 EHRs
  - [Li+20] learns to predict future diagnoses from 1.6 million EHRs

- **Text**

  [Dua+18] learns ICD-10 codes of cause of death from 121 536 free-text death certificates

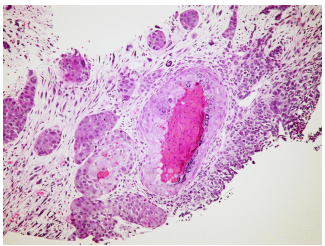
Deep learning is better suited to (big) non-tabular data

- **Images**
  [Gul+16] learns from 128 175 retinal fundus photographs
- **Graphs**
  [LU21] predicts chronic disease by creating a patient network from 1.2 million EHRs
- **Longitudinal data**
  - [Raj+18] learns medical events (unplanned readmission, mortality) from 216 221 EHRs
  - [Li+20] learns to predict future diagnoses from 1.6 million EHRs
- **Text**
  [Dua+18] learns ICD-10 codes of cause of death from 121 536 free-text death certificates

Deep learning for images

- One of the **most successful** areas of clinical applications of deep learning [Agg+21]
  - Detecting diabetic retinopathy [Gul+16]
  - Identifying skin disease [Est+17]

- **Machine learning for medical imaging: methodological failures and recommendations for the future** [VC22]



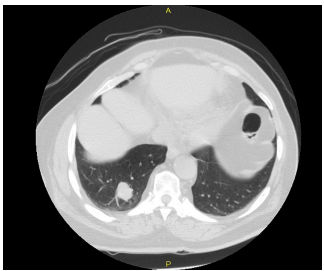


Image credits: Yale Rosen pulmonary_pathology (Flickr)

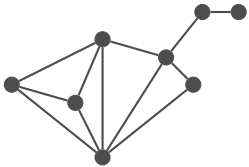
Deep learning is better suited to (big) non-tabular data

- **Images**
  [Gul+16] learns from 128 175 retinal fundus photographs
- **Graphs**
  [LU21] predicts chronic disease by creating a patient network from 1.2 million EHRs
- **Longitudinal data**
  - [Raj+18] learns medical events (unplanned readmission, mortality) from 216 221 EHRs
  - [Li+20] learns to predict future diagnoses from 1.6 million EHRs
- **Text**
  [Dua+18] learns ICD-10 codes of cause of death from 121 536 free-text death certificates

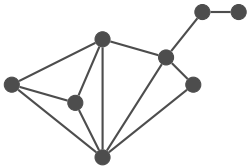
Deep learning for graphs

- Graphs model **relationships** between **entities:**

  genes, proteins, drugs, pathways, side effects, mutations, patients, lab tests, diagnoses, etc.



Deep learning for graphs



- Graphs model **relationships** between **entities:**

  genes, proteins, drugs, pathways, side effects, mutations, patients, lab tests, diagnoses, etc.

- **Learning on graphs:**
  - Label graphs
  - Label nodes
  - Label edges / edge completion

- Deep learning: **graph neural networks** [SL+21; Khe+24]; **graph transformers** [She+24]

Graph artificial intelligence in medicine [Joh+24]

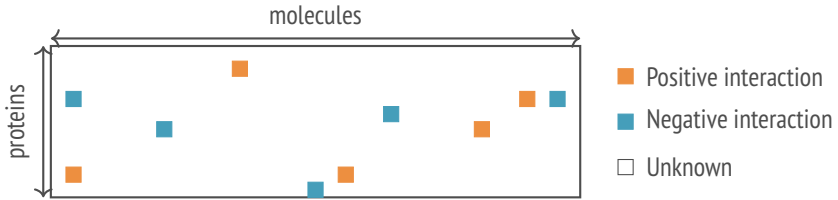
Tale 1: Large-scale drug-target interaction prediction

Gwenn Guichaoua et al., **Drug-target interaction prediction at scale: the Komet algorithm with the LCIdb dataset**, JCIM 2024 (read it on biorxiv)

`https://komet.readthedocs.io/`

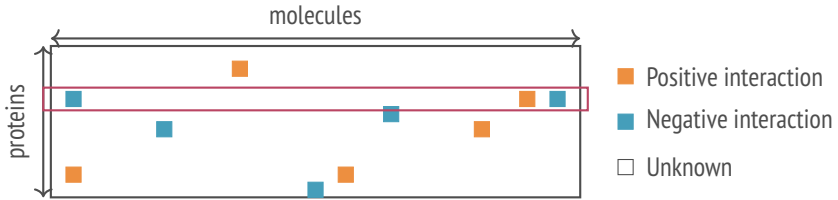
Large-scale drug-target interaction prediction

– **Goal**: predict whether a (protein, molecule) pair interacts



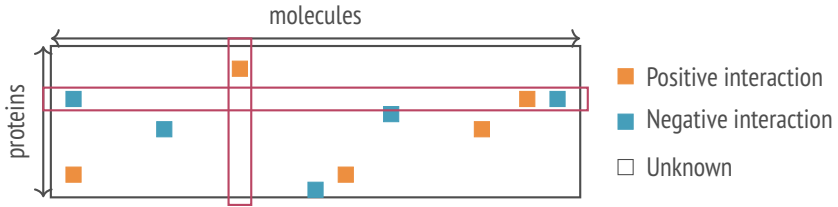
Large-scale drug-target interaction prediction

- **Goal**: predict whether a (protein, molecule) pair interacts
  - for a given protein and **all drug-like molecules** in a library: **drug discovery/repositioning**



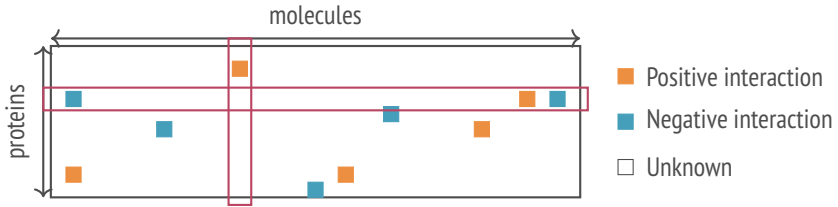
Large-scale drug-target interaction prediction

- **Goal**: predict whether a (protein, molecule) pair interacts
  - for a given protein and **all drug-like molecules** in a library: **drug discovery/repositioning**
  - for a given molecule and **all druggable proteins**: **side effects prediction**, **mechanism of action** of drugs identified by phenotypic screening



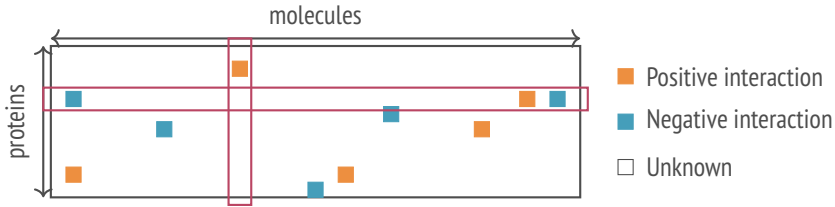
Large-scale drug-target interaction prediction

- **Goal**: predict whether a (protein, molecule) pair interacts
  - for a given protein and **all drug-like molecules** in a library: **drug discovery/repositioning**
  - for a given molecule and **all druggable proteins**: **side effects prediction**, **mechanism of action** of drugs identified by phenotypic screening
- **LCIdb**: 271 180 molecules, 2 060 proteins, 0.07% pairs labeled (98% positive)



Large-scale drug-target interaction prediction

- **Goal**: predict whether a (protein, molecule) pair interacts
  - for a given protein and **all drug-like molecules** in a library: **drug discovery/repositioning**
  - for a given molecule and **all druggable proteins**: **side effects prediction**, **mechanism of action** of drugs identified by phenotypic screening
- **LCIdb**: 271 180 molecules, 2 060 proteins, 0.07% pairs labeled (98% positive)
- Data is ☺ big and ☺ non-tabular; ☹ humans cannot perform the task



Large-scale drug-target interaction prediction

- **Methods**
  - **Molecule embedding** from molecular graph $\varphi_m : \mathfrak{m} \mapsto \vec{m} \in \mathbb{R}^{d_m}$
  - **Protein embedding** from protein sequence $\varphi_p : \mathfrak{p} \mapsto \vec{p} \in \mathbb{R}^{d_p}$
  - **Interaction model** $f : \mathbb{R}^{d_m} \times \mathbb{R}^{d_p} \to [0, 1]$

Large-scale drug-target interaction prediction

- **Methods**
  - **Molecule embedding** from molecular graph $\varphi_m : \mathfrak{m} \mapsto \vec{m} \in \mathbb{R}^{d_m}$
  - **Protein embedding** from protein sequence $\varphi_p : \mathfrak{p} \mapsto \vec{p} \in \mathbb{R}^{d_p}$
  - **Interaction model** $f : \mathbb{R}^{d_m} \times \mathbb{R}^{d_p} \to [0, 1]$

- State-of-the-art deep learning approaches[1]:

Ingredient	MolTrans [Hua+21]	ConPLex [Sin+23]
Molecule embedding	Transformers	ECFP
Protein embedding	Transformers	ProtBERT
Interaction model	neural network	neural network

[1] Now possibly all outperformed by CoNCISE [Erd+25]

Large-scale drug-target interaction prediction

- **Methods**
  - **Molecule embedding** from molecular graph $\varphi_m : \mathfrak{m} \mapsto \vec{m} \in \mathbb{R}^{d_m}$
  - **Protein embedding** from protein sequence $\varphi_p : \mathfrak{p} \mapsto \vec{p} \in \mathbb{R}^{d_p}$
  - **Interaction model** $f : \mathbb{R}^{d_m} \times \mathbb{R}^{d_p} \to [0, 1]$

Large-scale drug-target interaction prediction

- **Methods**
  - **Molecule embedding** from molecular graph $\varphi_m : \mathfrak{m} \mapsto \vec{m} \in \mathbb{R}^{d_m}$
  - **Protein embedding** from protein sequence $\varphi_p : \mathfrak{p} \mapsto \vec{p} \in \mathbb{R}^{d_p}$
  - **Interaction model** $f : \mathbb{R}^{d_m} \times \mathbb{R}^{d_p} \to [0, 1]$

- **Komet** (Kronecker Optimized METhod):
  - **Interaction model:** SVM with a Kronecker kernel

$$k((\mathfrak{m}, \mathfrak{p}), (\mathfrak{m}', \mathfrak{p}')) = k_m(\mathfrak{m}, \mathfrak{m}')k_p(\mathfrak{p}, \mathfrak{p}')$$

Large-scale drug-target interaction prediction

- **Methods**
  - **Molecule embedding** from molecular graph $\varphi_m : \mathfrak{m} \mapsto \vec{m} \in \mathbb{R}^{d_m}$
  - **Protein embedding** from protein sequence $\varphi_p : \mathfrak{p} \mapsto \vec{p} \in \mathbb{R}^{d_p}$
  - **Interaction model** $f : \mathbb{R}^{d_m} \times \mathbb{R}^{d_p} \to [0, 1]$

- **Komet** (Kronecker Optimized METhod):
  - **Interaction model:** SVM with a Kronecker kernel
    $$k((\mathfrak{m}, \mathfrak{p}), (\mathfrak{m}', \mathfrak{p}')) = k_m(\mathfrak{m}, \mathfrak{m}')k_p(\mathfrak{p}, \mathfrak{p}')$$
  - **Molecule embedding:** feature map corresponding to the molecule kernel k_m
    i.e. φ_m such that $k_m(\mathfrak{m}, \mathfrak{m}') \cong \langle \varphi_m(\mathfrak{m}), \varphi_m(\mathfrak{m}') \rangle$
  - **Protein embedding:** same

Large-scale drug-target interaction prediction

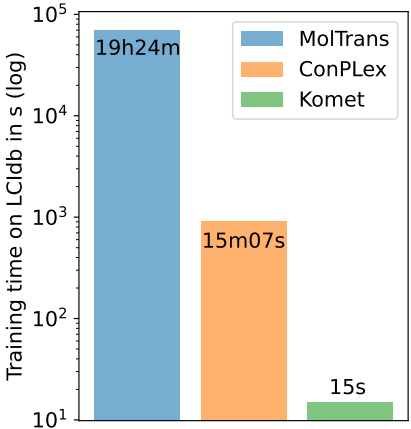
- **Methods**

Ingredient	MolTrans [Hua+21]	ConPLex [Sin+23]	**Komet**
Molecule embedding	transformers	ECFP	feature map of Tanimoto kernel on ECFP
Protein embedding	transformers	ProtBERT	feature map of Local Alignment kernel
Interaction model	neural network	neural network	SVM with Kronecker kernel

- **Komet** contribution: use Nyström approximations to **scale** to LCIdb

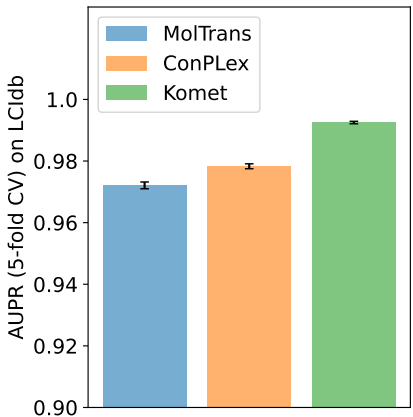
Large-scale drug-target interaction prediction

☺ Komet is **much faster** to train than its deep learning counterparts



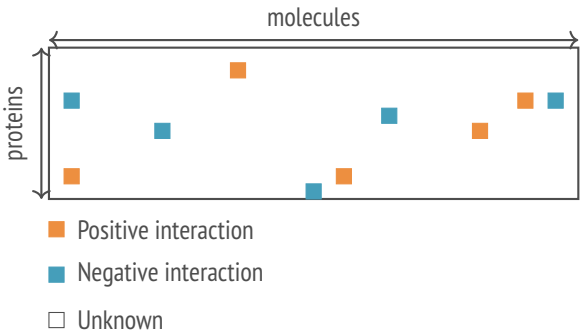
Large-scale drug-target interaction prediction

☺ Komet has **better performance** than its deep learning counterparts



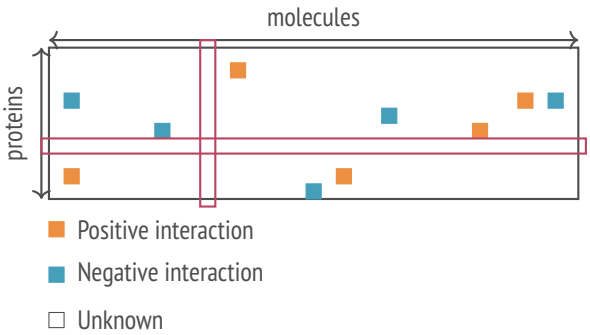
Large-scale drug-target interaction prediction

- ☺ Komet has **better performance** than its deep learning counterparts
- ☺ especially in the more difficult **orphan** setup



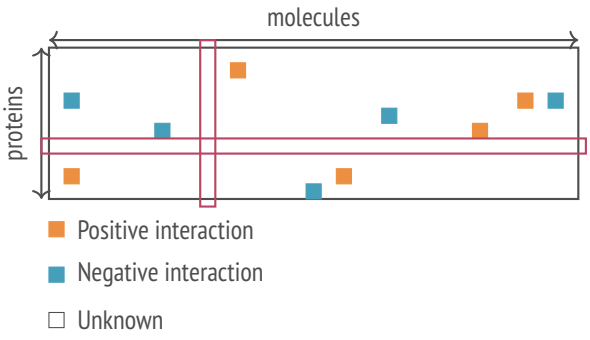
Large-scale drug-target interaction prediction

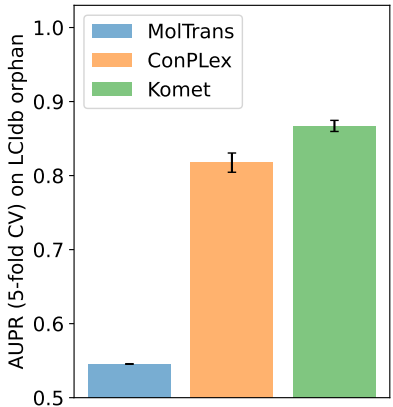
- ☺ Komet has **better performance** than its deep learning counterparts
- ☺ especially in the more difficult **orphan** setup



Large-scale drug-target interaction prediction

- ☺ Komet has **better performance** than its deep learning counterparts
- ☺ especially in the more difficult **orphan** setup





Deep learning is better suited to (big) non-tabular data

- **Images**
  [Gul+16] learns from 128 175 retinal fundus photographs
- **Graphs**
  [LU21] predicts chronic disease by creating a patient network from 1.2 million EHRs
- **Longitudinal data**
  - [Raj+18] learns medical events (unplanned readmission, mortality) from 216 221 EHRs
  - [Li+20] learns to predict future diagnoses from 1.6 million EHRs
- **Text**
  [Dua+18] learns ICD-10 codes of cause of death from 121 536 free-text death certificates

Deep learning for longitudinal data

– **Patient trajectory:** patient data acquired **through time**, at **irregular intervals**

diagnoses, therapies, lab values, clinical notes, imaging, etc.

Deep learning for longitudinal data

- **Patient trajectory:** patient data acquired **through time**, at **irregular intervals**
  diagnoses, therapies, lab values, clinical notes, imaging, etc.
- **LSTM** approaches
  - **LSTM** and **Bi-LSTM** networks: for time series (= regularly time-spaced)
  - **T-LSTM** [Bay+17]: also embeds time elapsed between consecutive elements
- **Transformer-based** approaches:
  - BEHRT [Li+20] predicts next diagnosis from sequence of previous diagnoses
  - [Moo+23] predicts risk of sepsis at ICU admissions using transformers

Machine and deep learning for longitudinal biomedical data [Cas+23]

Tale 2: Labeling relapse in breast cancer patient trajectories

Élise Dumas et al., **EDEN : An Event DEtection Network for the annotation of Breast Cancer recurrences in administrative claims data**, ML4Health 2022

`https://github.com/rt2lab/eden`

Event annotation in patient trajectories

- **Goal:** Identify **which patients** underwent **relapse**, and **when**
- **Data:**
  - 5 892 patients treated for primary breast cancer
  - **irregularly** spaced, **right-censored** sequences of **medical codes** from healthcare claims

Event annotation in patient trajectories

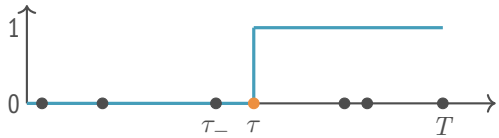
- **Goal:** Identify **which patients** underwent **relapse**, and **when**
- **Data:**
  - 5 892 patients treated for primary breast cancer
  - **irregularly** spaced, **right-censored** sequences of **medical codes** from healthcare claims
- Data is ☹ not so big and ☺ non-tabular; ☺ humans can (more or less) perform the task

Event annotation in patient trajectories

- **Goal:** Identify **which patients** underwent **relapse**, and **when**
- **Data:**
  - 5 892 patients treated for primary breast cancer
  - **irregularly** spaced, **right-censored** sequences of **medical codes** from healthcare claims
- Data is ☹ not so big and ☺ non-tabular; ☺ humans can (more or less) perform the task
- **Methods**
  - **T-LSTM** [Bay+17] does not deal with censorship

Event annotation in patient trajectories

- **EDEN:** Extends **T-LSTM** to **censored data**, using a new **loss function** to estimate the **event rate**
- **Patient event rate function:** $W(t) = \mathbb{P}[t \geq \tau]$ τ = time to event



- **Loss function:**

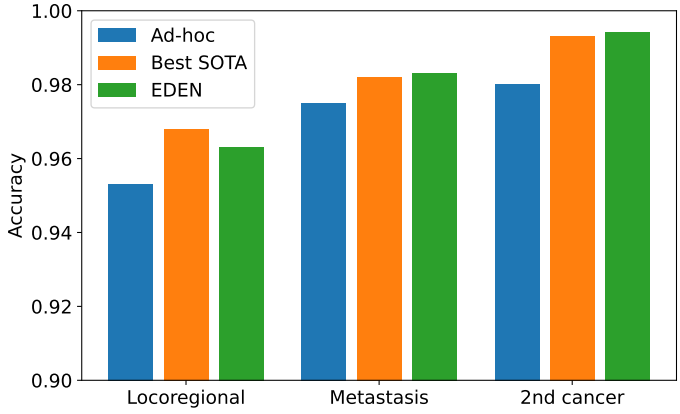
$$\underbrace{\alpha_1 \mathcal{L}_1}_{\text{BCE}} + \underbrace{\alpha_2 \mathcal{L}_2 + \alpha_3 \mathcal{L}_3}_{\text{uncensored}} + \underbrace{\alpha_4 \mathcal{L}_4}_{\text{censored}}$$

  - **uncensored data:** $\mathcal{L}_2$ penalizes low $\widetilde{W}(\tau)$ and $\mathcal{L}_3$ penalizes high $\widetilde{W}(\tau_-)$
  - **censored data:** $\mathcal{L}_4$ penalizes high $\widetilde{W}(T)$

Event annotation in patient trajectories

Performance on an **independent test set** of 800 patient trajectories

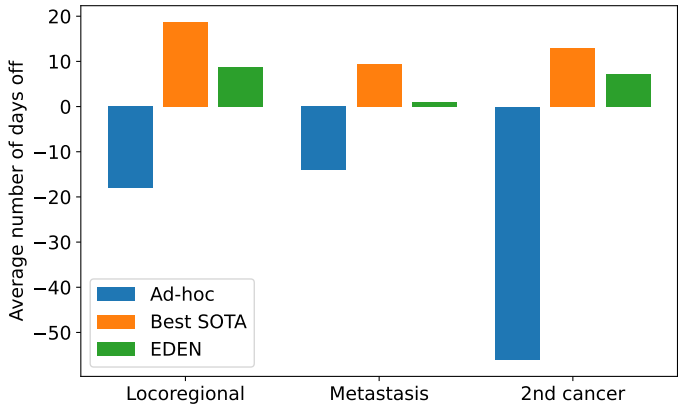
☺ EDEN achieves **state-of-the-art accuracy** in predicting **whether** there is a relapse



Event annotation in patient trajectories

Performance on an **independent test set** of 800 patient trajectories

☺ EDEN is much better than other approaches at predicting **when** the relapse occurred



Event annotation in patient trajectories

- ☺ EDEN achieves **state-of-the-art accuracy** in predicting **whether** there is a relapse
- ☺ EDEN is much better than other approaches at predicting **when** the relapse occurred

Event annotation in patient trajectories

- ☺ EDEN achieves **state-of-the-art accuracy** in predicting **whether** there is a relapse
- ☺ EDEN is much better than other approaches at predicting **when** the relapse occurred

However these results

- ☹ Only show **one performance criterion** (accuracy)
- ☹ Are presented **without confidence intervals**

Event annotation in patient trajectories

☺ EDEN achieves **state-of-the-art accuracy** in predicting **whether** there is a relapse

☺ EDEN is much better than other approaches at predicting **when** the relapse occurred

However these results

☹ Only show **one performance criterion** (accuracy)

☹ Are presented **without confidence intervals**

☹ Do not **improve enough** over the ad-hoc approach that we are ready to use EDEN in practice.

Tale 3: Predicting outcome from multi-modal patient trajectories

Maguette Ndèye Mbaye et al., **Multimodal BEHRT: transformers for multimodal Electronic Health Records**, medrxiv 2024

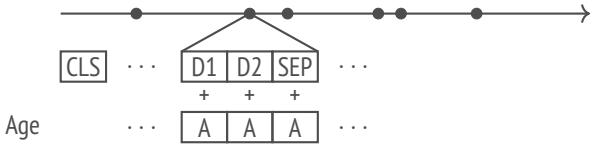
`https://github.com/maguettemb/Multimodal-BEHRT`

Classification of multi-modal patient trajectories

- **Goal:** Predict disease-free survival status 3 years after first breast cancer surgery
- **Data:**
  - 15 150 patients treated with adjuvant chemotherapy
  - **multi-modal trajectories:** rich tabular information for each visit until 1 year after surgery
    - biological measurements
    - clinical information (age, tumor size/grade/subtype, number of lymph nodes involved)
    - treatment
    - department/procedure
- Data is ☹ not so big and ☺ non-tabular; ☺ humans cannot perform the task

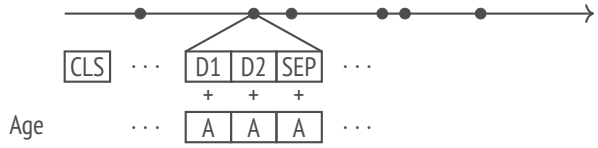
Classification of multi-modal patient trajectories

– **BEHRT** [Li+20] uses one type of information per medical event

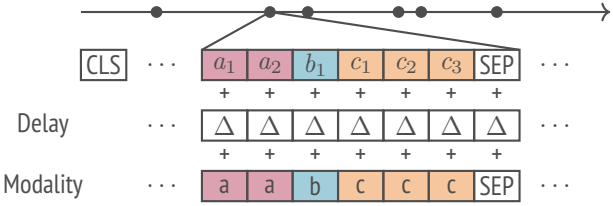


Classification of multi-modal patient trajectories

- **BEHRT** [Li+20] uses one type of information per medical event



- **TabularBEHRT**

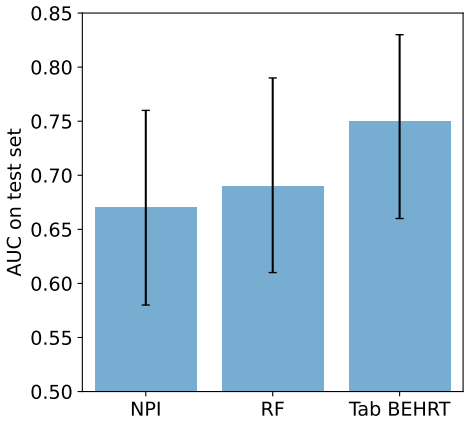


Classification of multi-modal patient trajectories

- **Nottingham Prognostic Index (NPI):** $0.2T + G + N$

  T = tumor size (cm); G = tumor grade; N = 1 if 0 lymph node involved, 2 if 1 to 3 nodes, 3 otherwise

☹ **TabularBEHRT** does not significantly outperform **NPI** on held-out data



Deep learning is better suited to (big) non-tabular data

- **Images**

  [Gul+16] learns from 128 175 retinal fundus photographs

- **Graphs**

  [LU21] predicts chronic disease by creating a patient network from 1.2 million EHRs

- **Longitudinal data**
  - [Raj+18] learns medical events (unplanned readmission, mortality) from 216 221 EHRs
  - [Li+20] learns to predict future diagnoses from 1.6 million EHRs

- **Text**

  [Dua+18] learns ICD-10 codes of cause of death from 121 536 free-text death certificates

Deep learning for text

– Lots of **free text** clinical notes and reports

[2] nowadays Recurrent Memory Transformers can deal with 1 million tokens [Bul+24].

Deep learning for text

- Lots of **free text** clinical notes and reports
- Challenges:
  - **Highly specialized** text (specific vocabulary; acronyms and shorthand)
  - Language is **not necessarily English**

[2] nowadays Recurrent Memory Transformers can deal with 1 million tokens [Bul+24].

Deep learning for text

- Lots of **free text** clinical notes and reports
- Challenges:
  - **Highly specialized** text (specific vocabulary; acronyms and shorthand)
  - Language is **not necessarily English**

⇒ **Specialized models** are better than generic ones [Leh+23]

  - Models trained on US clinical notes: ClinicalBERT [Als+19], GatorTron [Yan+22]
  - DrBERT [Lab+23] trained on French clinical notes

[2] nowadays Recurrent Memory Transformers can deal with 1 million tokens [Bul+24].

Deep learning for text

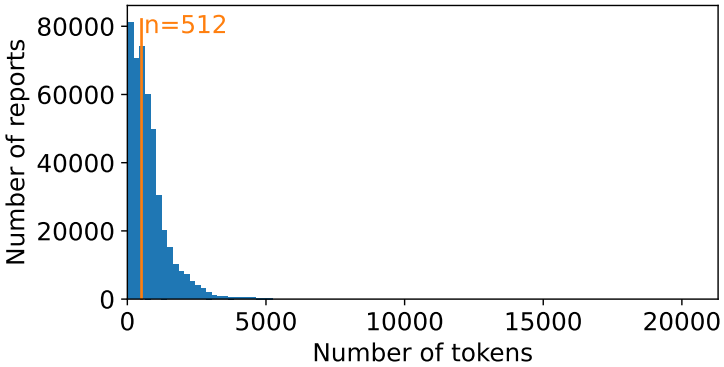
- Lots of **free text** clinical notes and reports
- Challenges:
  - **Highly specialized** text (specific vocabulary; acronyms and shorthand)
  - Language is **not necessarily English**

  ⇒ **Specialized models** are better than generic ones [Leh+23]
    - Models trained on US clinical notes: ClinicalBERT [Als+19], GatorTron [Yan+22]
    - DrBERT [Lab+23] trained on French clinical notes
  - **Long** text but BERT is limited to 512 tokens (self-attention)[2]

[2] nowadays Recurrent Memory Transformers can deal with 1 million tokens [Bul+24].

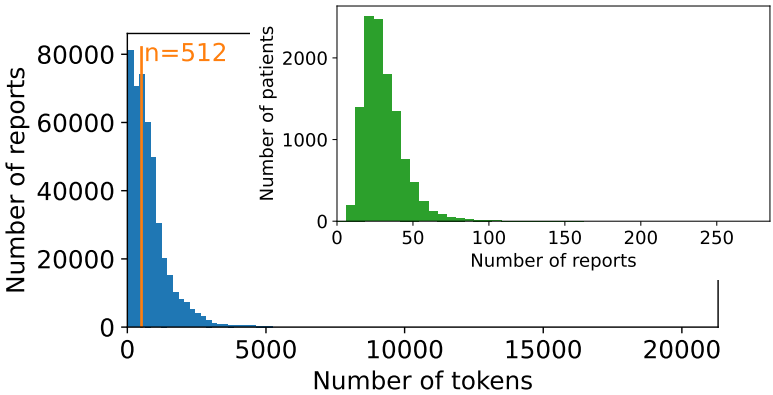
Text BEHRT

- **Data:**
  - 15 150 patients treated with adjuvant chemotherapy
  - **text-based trajectories:** free-text medical reports for each visit



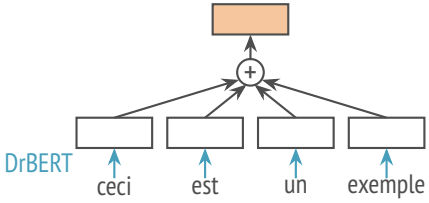
Text BEHRT

- **Data:**
  - 15 150 patients treated with adjuvant chemotherapy
  - **text-based trajectories:** free-text medical reports for each visit



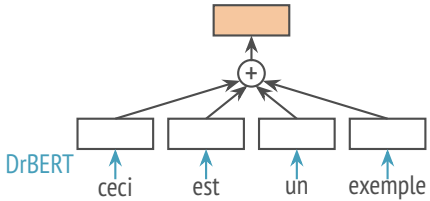
Text BEHRT

– **Report embedding:** pool token embeddings obtained by DrBERT [Lab+23]

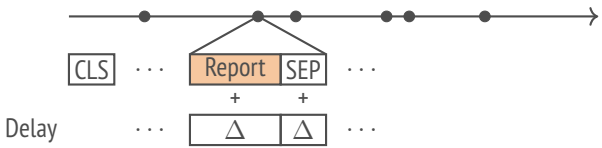


Text BEHRT

– **Report embedding:** pool token embeddings obtained by DrBERT [Lab+23]



– **TextBEHRT:** Use a BEHRT-inspired architecture and embedding again



Text BEHRT

- **Nottingham Prognostic Index (NPI):** $0.2T + G + N$

  T = tumor size (cm); G = tumor grade; N = 1 if 0 lymph node involved, 2 if 1 to 3 nodes, 3 otherwise

☹ **TabularBEHRT** does not significantly outperform **NPI** on held-out data

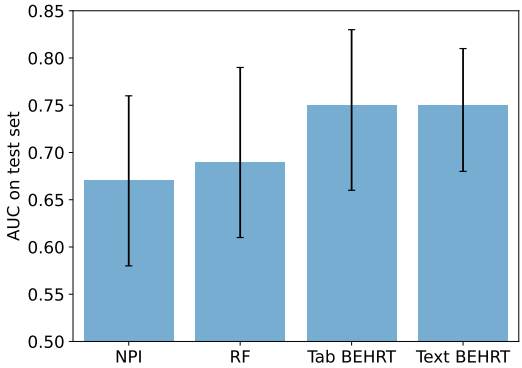
Text BEHRT

- **Nottingham Prognostic Index (NPI):** $0.2T + G + N$

  T = tumor size (cm); G = tumor grade; N = 1 if 0 lymph node involved, 2 if 1 to 3 nodes, 3 otherwise

☹ **TabularBEHRT** does not significantly outperform **NPI** on held-out data

☹ and neither does **Text BEHRT**



Multimodal BEHRT

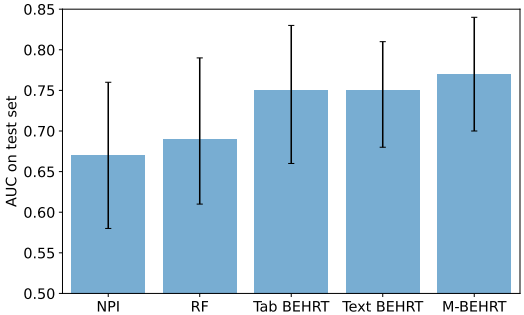
- **M-BEHRT: Combine** Tabular BEHRT and Text BEHRT through **cross-attention** [CFP21]

Multimodal BEHRT

- **M-BEHRT: Combine** Tabular BEHRT and Text BEHRT through **cross-attention** [CFP21]

☺ **M-BEHRT** significantly outperforms **NPI** and **random forests** on held-out data

☺ **Interpretation** through **integrated gradients** [STY17] pointed towards possible prognostic factors



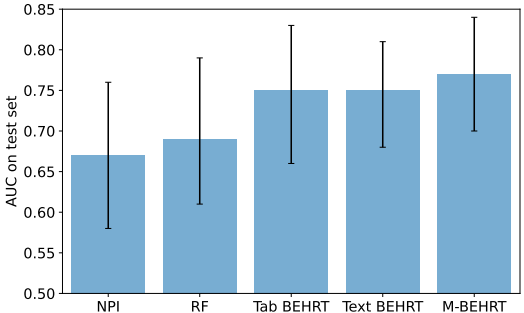
Multimodal BEHRT

- **M-BEHRT: Combine** Tabular BEHRT and Text BEHRT through **cross-attention** [CFP21]

☺ **M-BEHRT** significantly outperforms **NPI** and **random forests** on held-out data

☺ **Interpretation** through **integrated gradients** [STY17] pointed towards possible prognostic factors

☹ Still a long way to go for **clinical utility**



Take-home messages

- Recent advances in deep learning are **exciting**, but **proceed with caution**

Take-home messages

- Recent advances in deep learning are **exciting**, but **proceed with caution**
- **Data** used in the clinic is often
  - small
  - tabular and/or multimodal
  - not so well understood in nature

Take-home messages

- Recent advances in deep learning are **exciting**, but **proceed with caution**
- **Data** used in the clinic is often
  - small
  - tabular and/or multimodal
  - not so well understood in nature
- **Gains in performance** are often not enough
  - Is the **performance metric** appropriate?
  - Is the gain **significant?**
  - Are the **resources** (time, computational cost, ecological footprint) involved worth the gain?
  - **Interpretability** & **trustworthiness** are key

Thanks

CBIO: **Élise Dumas** (now EPFL), **Gwenn Guichaoua**, **Maguette Ndèye Mbaye**, Philippe Pinel (now Iktos), Véronique Stoven.

Institut Curie: Beatriz Grandal, Julien Guérin, Anne-Sophie Hamy-Petit, Fabien Reyal, Aullène Toussaint.

IBM Research: Michael Danziger, Michal Rosen-Zvi.

CLINDA Aalborg: Martin Bøgsted, Charles Vesteghem.

Funding: European Research Council (MLFPM2018, H2020-MSCA-ITN-2018 813533); Agence Nationale pour la Recherche (PRAIRIE); Paris Île-de-France (DIM AI4IDF).

References I

[Agg+21] Ravi Aggarwal et al. "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis". In: *npj Digital Medicine* 4.1 (2021). Publisher: Nature Publishing Group, pp. 1–23. ISSN: 2398-6352. DOI: 10.1038/s41746-021-00438-z.

[Als+19] Emily Alsentzer et al. "Publicly Available Clinical BERT Embeddings". In: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Ed. by Anna Rumshisky et al. Minneapolis, Minnesota, USA: Association for Computational Linguistics, 2019, pp. 72–78. DOI: 10.18653/v1/W19-1909.

[Bay+17] Inci M. Baytas et al. "Patient Subtyping via Time-Aware LSTM Networks". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 65–74. ISBN: 978-1-4503-4887-4. DOI: 10.1145/3097983.3097997.

[Bul+24] Aydar Bulatov et al. *Scaling Transformer to 1M tokens and beyond with RMT*. 2024. DOI: 10.48550/arXiv.2304.11062. URL: http://arxiv.org/abs/2304.11062.

[Cas+23] Anna Cascarano et al. "Machine and deep learning for longitudinal biomedical data: a review of methods and applications". In: *Artificial Intelligence Review* 56.S2 (2023), pp. 1711–1771. ISSN: 0269-2821, 1573-7462. DOI: 10.1007/s10462-023-10561-w.

[CFP21] Chun-Fu Richard Chen, Quanfu Fan, and Rameswar Panda. "Crossvit: Cross-attention multi-scale vision transformer for image classification". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pp. 357–366.

References II

[Dua+18] Francisco Duarte et al. "Deep neural models for ICD-10 coding of death certificates and autopsy reports in free-text". In: *Journal of Biomedical Informatics* 80 (2018), pp. 64–77. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2018.02.011.

[Dum+22] Elise Dumas et al. *EDEN : An Event DEtection Network for the annotation of Breast Cancer recurrences in administrative claims data*. arXiv:2211.08077 [cs]. 2022. DOI: 10.48550/arXiv.2211.08077.

[Erd+25] Mert Erden et al. "Learning aăCoNCISE Language forăSmall-Molecule Binding". In: *Research in Computational Molecular Biology: 29th International Conference, RECOMB 2025, Seoul, South Korea, April 2629, 2025, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 2025, pp. 260–263. ISBN: 978-3-031-90251-2. DOI: 10.1007/978-3-031-90252-9_17. URL: https://doi.org/10.1007/978-3-031-90252-9_17.

[Est+17] Andre Esteva et al. "Dermatologist-level classification of skin cancer with deep neural networks". In: *Nature* 542.7639 (2017). Publisher: Nature Publishing Group, pp. 115–118. ISSN: 1476-4687. DOI: 10.1038/nature21056.

[GOV22] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. "Why do tree-based models still outperform deep learning on typical tabular data?" In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. NIPS '22. Red Hook, NY, USA: Curran Associates Inc., 2022, pp. 507–520. ISBN: 978-1-71387-108-8.

[Gui+24] Gwenn Guichaoua et al. *Drug-Target Interactions Prediction at Scale: the Komet Algorithm with the LCIdb Dataset*. Pages: 2024.02.22.581599 Section: New Results. 2024. DOI: 10.1101/2024.02.22.581599.

References III

[Gul+16] Varun Gulshan et al. "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs". In: *JAMA* 316.22 (2016), pp. 2402–2410. ISSN: 1538-3598. DOI: 10.1001/jama.2016.17216.

[Hol+25] Noah Hollmann et al. "Accurate predictions on small data with a tabular foundation model". In: *Nature* 637.8045 (2025), pp. 319–326. ISSN: 1476-4687. DOI: 10.1038/s41586-024-08328-6. URL: https://www.nature.com/articles/s41586-024-08328-6 (visited on 05/19/2025).

[Hua+21] Kexin Huang et al. "MolTrans: Molecular Interaction Transformer for drugtarget interaction prediction". In: *Bioinformatics* 37.6 (2021), pp. 830–836. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btaa880.

[Joh+24] Ruth Johnson et al. "Graph Artificial Intelligence in Medicine". In: *Annual Review of Biomedical Data Science* 7.Volume 7, 2024 (2024). Publisher: Annual Reviews, pp. 345–368. ISSN: 2574-3414. DOI: 10.1146/annurev-biodatasci-110723-024625.

[KGV24] Myung Jun Kim, Léo Grinsztajn, and Gaël Varoquaux. *CARTE: Pretraining and Transfer for Tabular Learning.* arXiv:2402.16785 [cs]. 2024. DOI: 10.48550/arXiv.2402.16785.

[Khe+24] Bharti Khemani et al. "A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions". In: *Journal of Big Data* 11.1 (2024). Number: 1 Publisher: SpringerOpen, pp. 1–43. ISSN: 2196-1115. DOI: 10.1186/s40537-023-00876-4.

References IV

[Lab+23] Yanis Labrak et al. "DrBERT: A Robust Pre-trained Model in French for Biomedical and Clinical domains". In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki. Association for Computational Linguistics, 2023, pp. 16207–16221. DOI: 10.18653/v1/2023.acl-long.896.

[Leh+23] Eric Lehman et al. "Do We Still Need Clinical Language Models?" In: *Proceedings of the Conference on Health, Inference, and Learning*. ISSN: 2640-3498. PMLR, 2023, pp. 578–597.

[Li+20] Yikuan Li et al. "BEHRT: Transformer for Electronic Health Records". In: *Scientific Reports* 10.1 (2020). Number: 1 Publisher: Nature Publishing Group, p. 7155. ISSN: 2045-2322. DOI: 10.1038/s41598-020-62922-y.

[LU21] Haohui Lu and Shahadat Uddin. "A weighted patient network-based framework for predicting chronic diseases using graph neural networks". In: *Scientific Reports* 11.1 (Nov. 2021), pp. 1–12. ISSN: 2045-2322. DOI: 10.1038/s41598-021-01964-2. URL: https://www.nature.com/articles/s41598-021-01964-2.

[McE+23] Duncan McElfresh et al. "When do neural nets outperform boosted trees on tabular data?" In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. NIPS '23. Curran Associates Inc., 2023, pp. 76336–76369.

[Moo+23] Michael Moor et al. "Predicting sepsis using deep learning across international sites: a retrospective development and validation study". In: *eClinicalMedicine* 62 (2023). Publisher: Elsevier. ISSN: 2589-5370. DOI: 10.1016/j.eclinm.2023.102124.

References V

[Ng04] Andrew Y. Ng. "Feature selection, L1 vs. L2 regularization, and rotational invariance". In: *Proceedings of the twenty-first international conference on Machine learning*. ICML '04. New York, NY, USA: Association for Computing Machinery, 2004, p. 78. ISBN: 978-1-58113-838-2. DOI: 10.1145/1015330.1015435.

[Raj+18] Alvin Rajkomar et al. "Scalable and accurate deep learning with electronic health records". In: *npj Digital Medicine* 1.1 (2018). Publisher: Nature Publishing Group, pp. 1–10. ISSN: 2398-6352. DOI: 10.1038/s41746-018-0029-1.

[She+24] Ahsan Shehzad et al. *Graph Transformers: A Survey*. arXiv:2407.09777 [cs]. 2024. DOI: 10.48550/arXiv.2407.09777.

[Sin+23] Rohit Singh et al. "Contrastive learning in protein language space predicts interactions between drugs and protein targets". In: *Proceedings of the National Academy of Sciences of the United States of America* 120.24 (2023), e2220778120. ISSN: 1091-6490. DOI: 10.1073/pnas.2220778120.

[SL+21] Benjamin Sanchez-Lengeling et al. "A Gentle Introduction to Graph Neural Networks". In: *Distill* 6.9 (2021), e33. ISSN: 2476-0757. DOI: 10.23915/distill.00033.

[SP24] Carsen Stringer and Marius Pachitariu. *Transformers do not outperform Cellpose*. Pages: 2024.04.06.587952 Section: New Results. 2024. DOI: 10.1101/2024.04.06.587952.

[STY17] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. "Axiomatic attribution for deep networks". In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. ICML'17. Sydney, NSW, Australia: JMLR.org, 2017, pp. 33193328.

References VI

[Tou+23] Hugo Touvron et al. *Llama 2: Open Foundation and Fine-Tuned Chat Models*. arXiv:2307.09288 [cs]. 2023. DOI: 10.48550/arXiv.2307.09288.

[VC22] Gaël Varoquaux and Veronika Cheplygina. "Machine learning for medical imaging: methodological failures and recommendations for the future". en. In: *npj Digital Medicine* 5.1 (2022). Publisher: Nature Publishing Group, pp. 1–8. ISSN: 2398-6352. DOI: 10.1038/s41746-022-00592-y.

[Yan+22] Xi Yang et al. "A large language model for electronic health records". In: *npj Digital Medicine* 5.1 (2022). Publisher: Nature Publishing Group, pp. 1–9. ISSN: 2398-6352. DOI: 10.1038/s41746-022-00742-2.