

# Integrating complex multimodal health data for clinical prediction with artificial intelligence



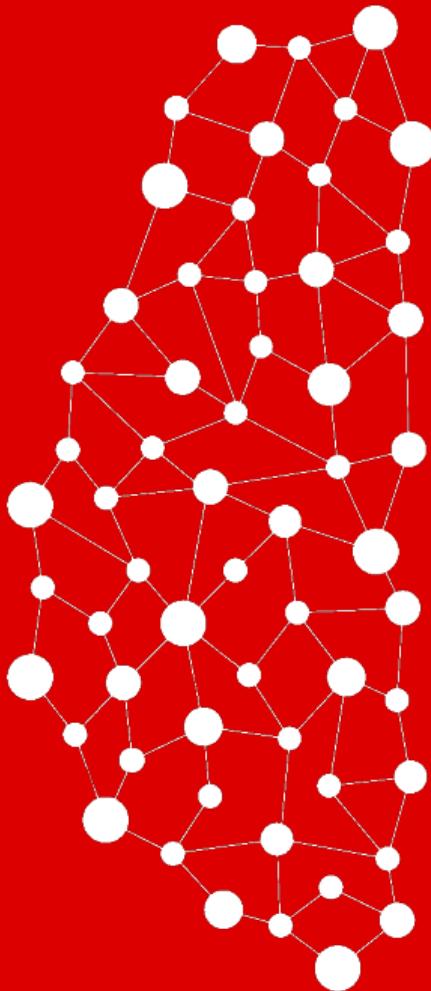
Esten H. Leonardsen

Post-doktor ved Psykologisk institutt,  
Universitet i Oslo

Chief Scientific Officer, baba.vision



UNIVERSITETET  
I OSLO



# What is artificial intelligence?

## What is AI?

Artificial Intelligence (AI) is the field of computer science devoted to creating systems that can perform tasks that typically require human intelligence. These tasks include perception (e.g., recognizing images or speech), reasoning (e.g., solving problems, making decisions), learning (e.g., adapting from data and experience), and interaction (e.g., understanding natural language, generating responses, collaborating with humans).

At a high level, AI can be divided into:

- Symbolic AI (knowledge-based, rule-based, expert systems): Uses explicit rules, logic, and symbolic representations to encode reasoning.
- Machine Learning (data-driven): Learns patterns from data, with subfields like deep learning, reinforcement learning, and probabilistic modeling.

AI is not a single method but a spectrum of approaches for replicating or augmenting cognitive functions in machines.

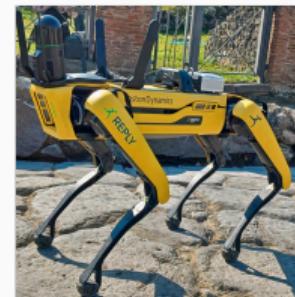
Would you like me to give you a concise working definition (one sentence) or a taxonomy of types of AI (narrow, general, superintelligence, etc.)?



+ Ask anything



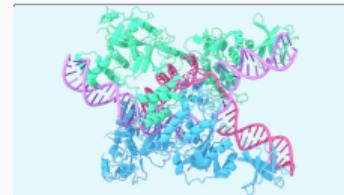
ChatGPT



Spot



Sophia

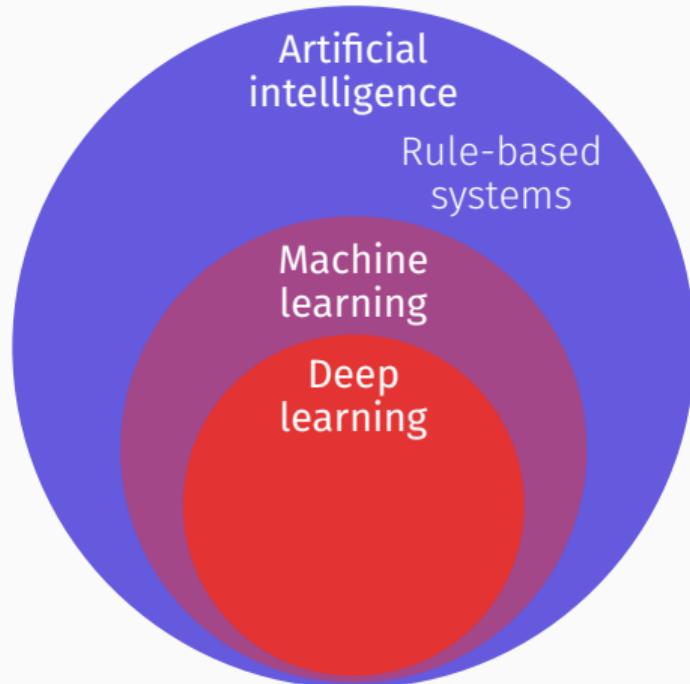


AlphaFold



AlphaZero

# What is artificial intelligence?



## Artificial intelligence

The field of study producing technology that solves tasks requiring some form of intelligence

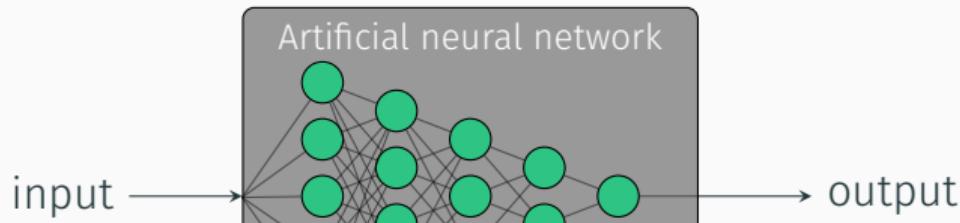
## Machine learning

A set of techniques to solve problems by allowing machines to find patterns in training data

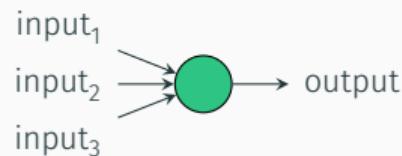
## Deep learning

Machine learning approaches that rely on artificial neural networks

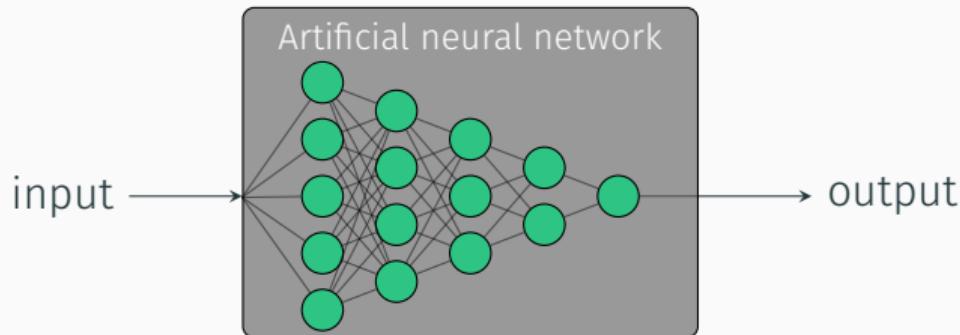
# Why do we use artificial neural networks?



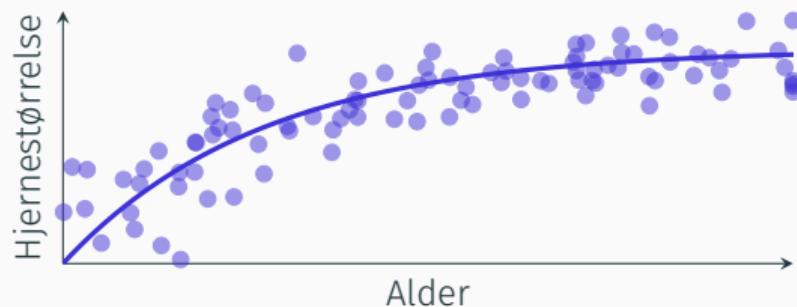
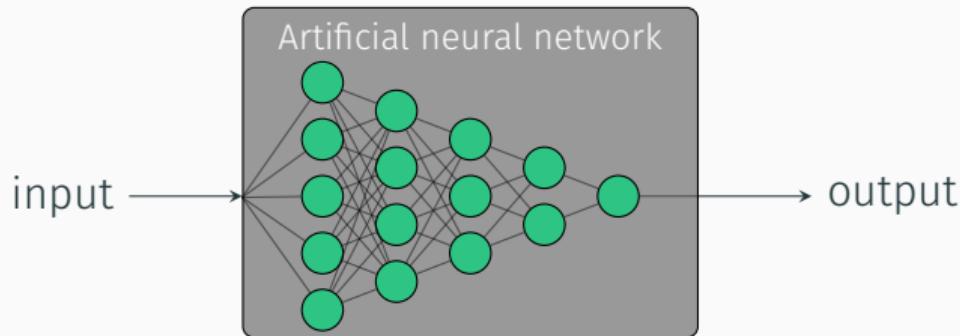
Artificial neuron



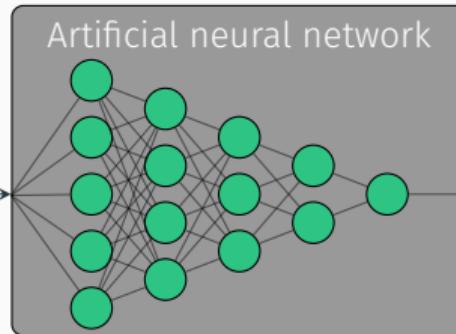
# Why do we use artificial neural networks?



# Why do we use artificial neural networks?



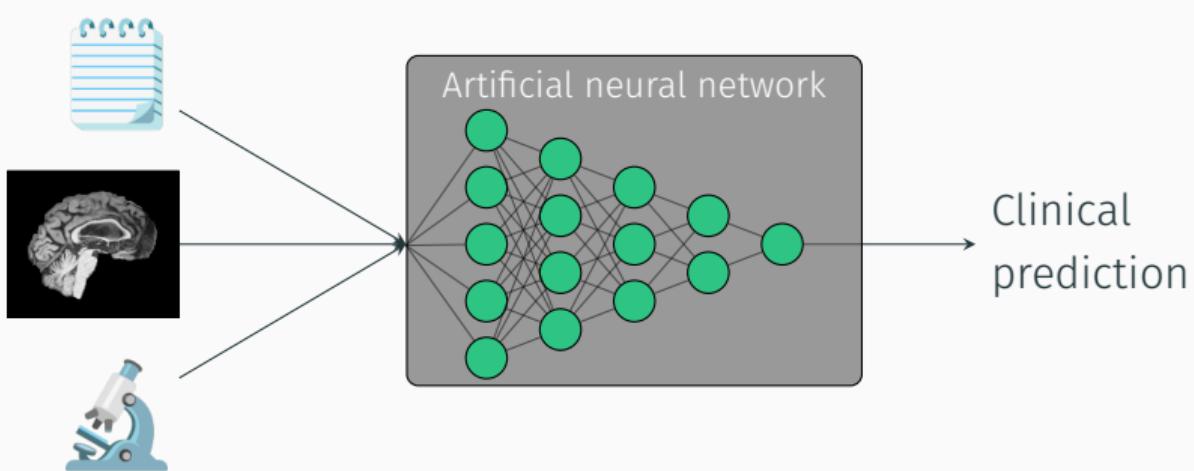
# Why do we use artificial neural networks?



ladybug



# Why do we use artificial neural networks?



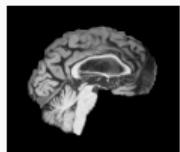
# Integrating multimodal health data for clinical predictions

---

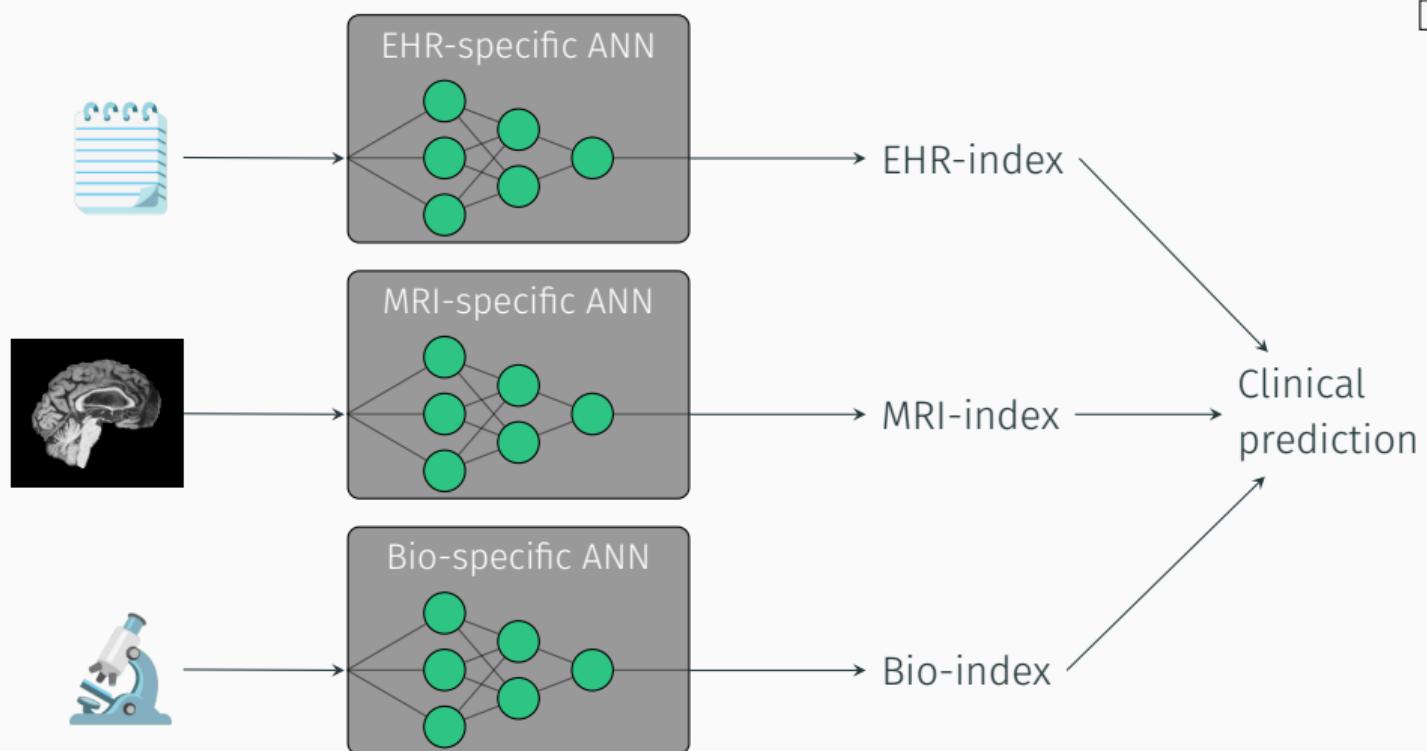


UNIVERSITETET  
I OSLO

# Late fusion: independent insights, combined decisions



# Late fusion: independent insights, combined decisions



# Late fusion: independent insights, combined decisions



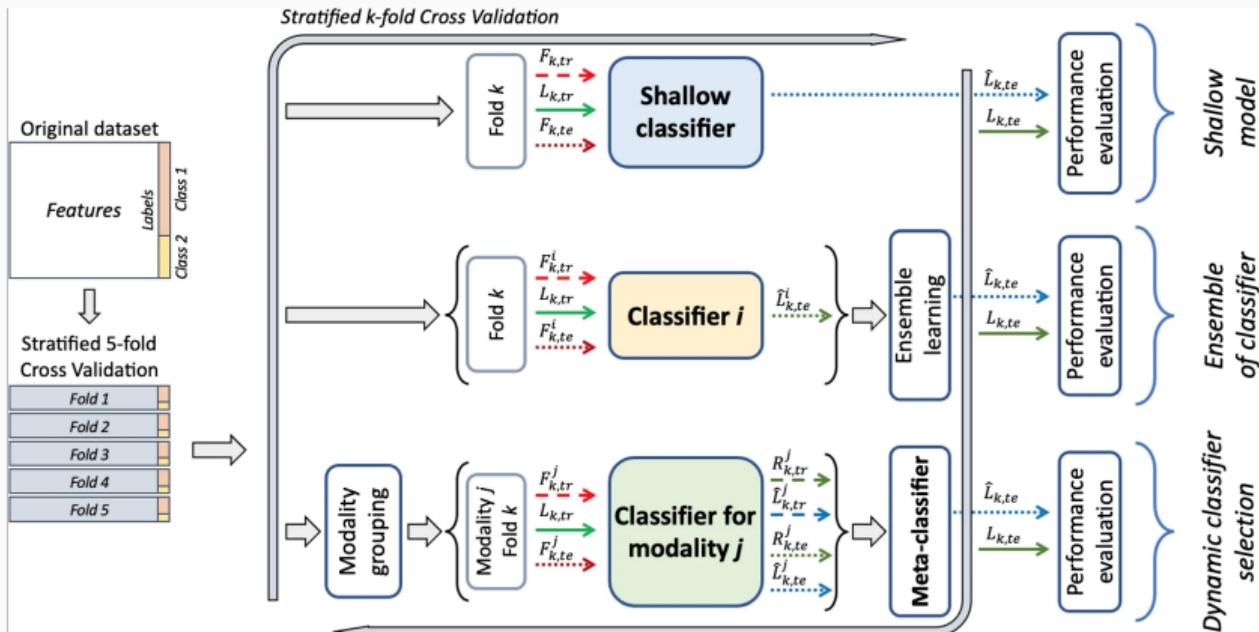
- Modality 1: Anamnestic data, which includes Age at hospitalization, Sex, Cardiovascular disorders, Neurological disorders, Oncological disorders, Diabetes, Obesity, Dyspnea and Smoking attitude;
- Modality 2: Hospitalisation data, which includes X-Ray at hospitalization, unenhanced chest CT at hospitalization, Days of fever at hospitalization, Lactate Dehydrogenase (LDH) at hospitalization, C Reactive Protein (CRP) at hospitalization, D-dimer at hospitalization, Creatinine at hospitalization, X-Ray at discharge, unenhanced chest CT at discharge, CRP at discharge and Creatinine at discharge;
- Modality 3: Ventilation data, which includes Peripheral Oxygen Saturation ( $\text{SpO}_2$ ) at hospitalization, Oxygen Saturation ( $\text{O}_2$ ) at hospitalization, Days of ventilation, Continuous positive airway pressure (CPAP) ventilation and Non-invasive ventilation (NIV) ventilation.



Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).



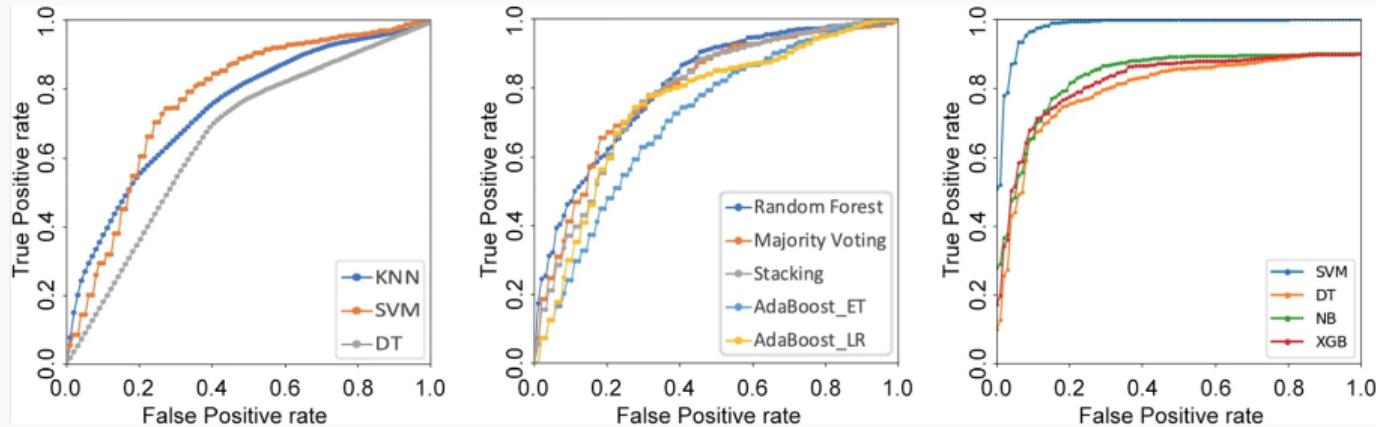
# Late fusion: independent insights, combined decisions



Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).



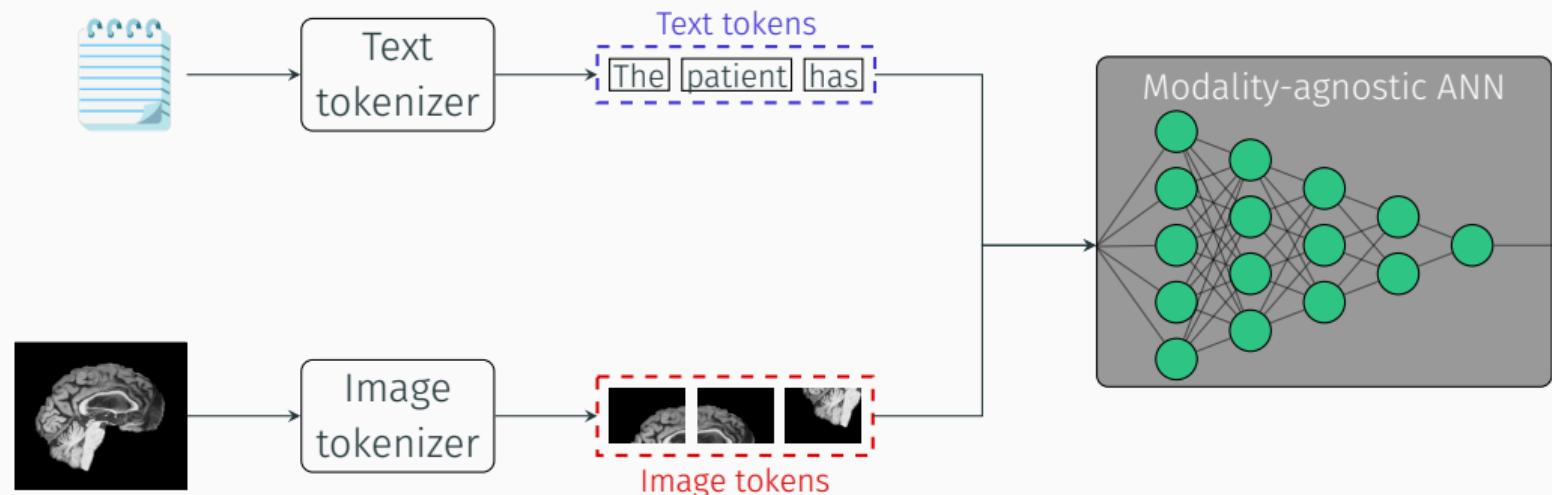
# Late fusion: independent insights, combined decisions



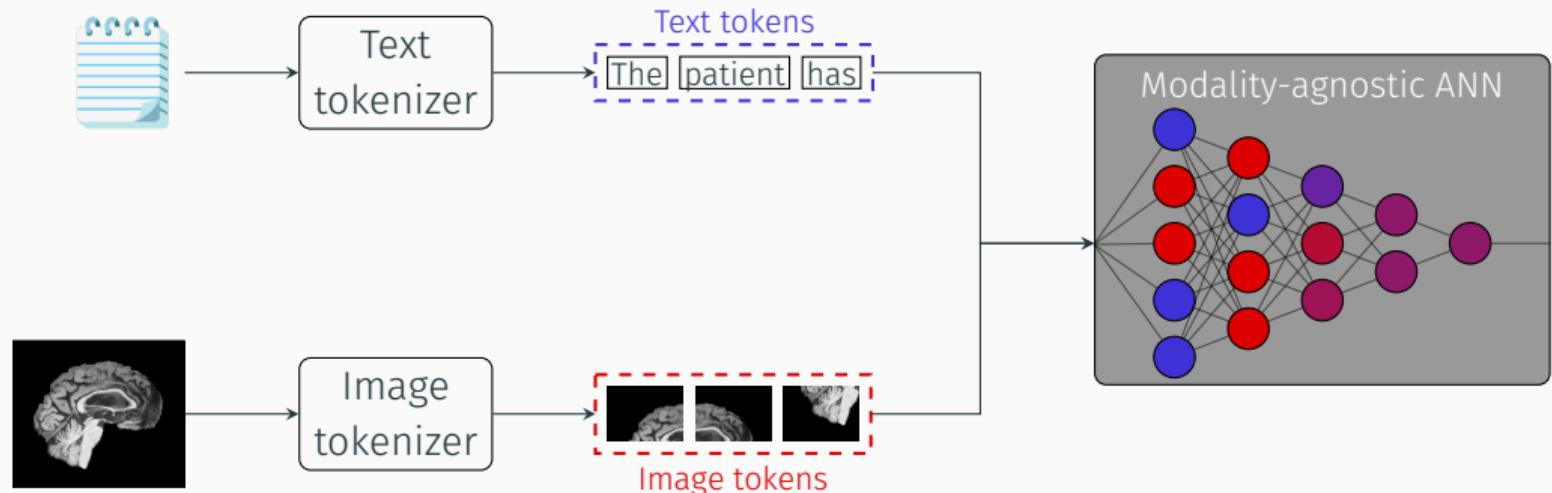
Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).



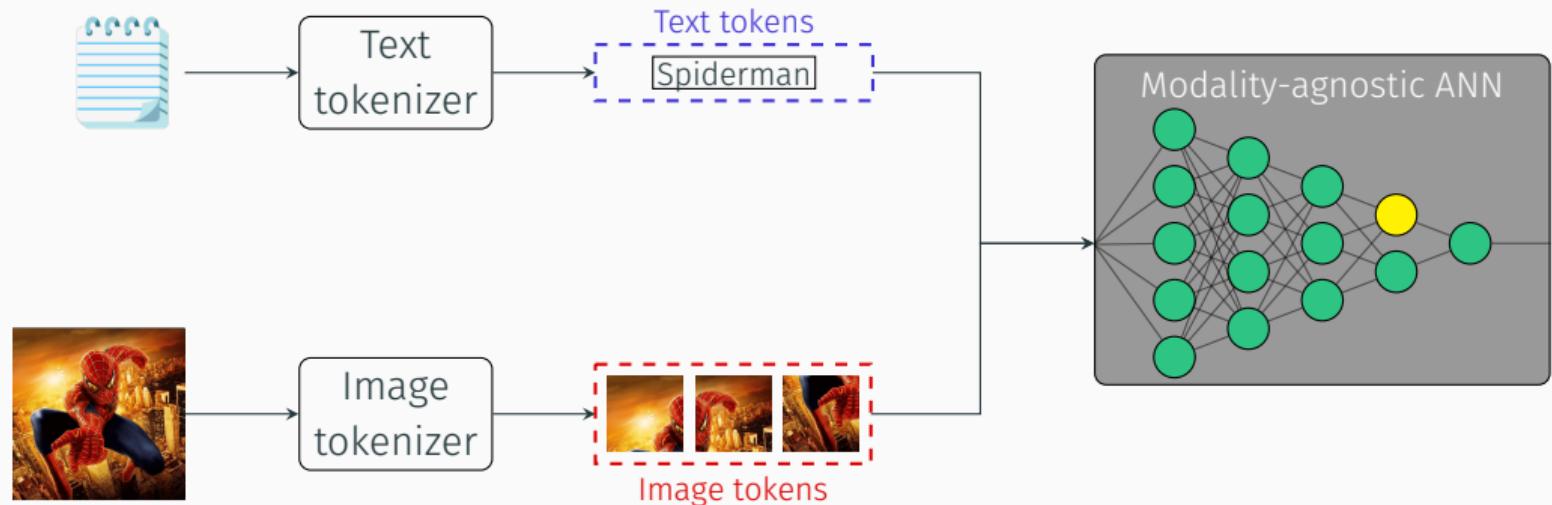
# Early fusion: blending information from the start



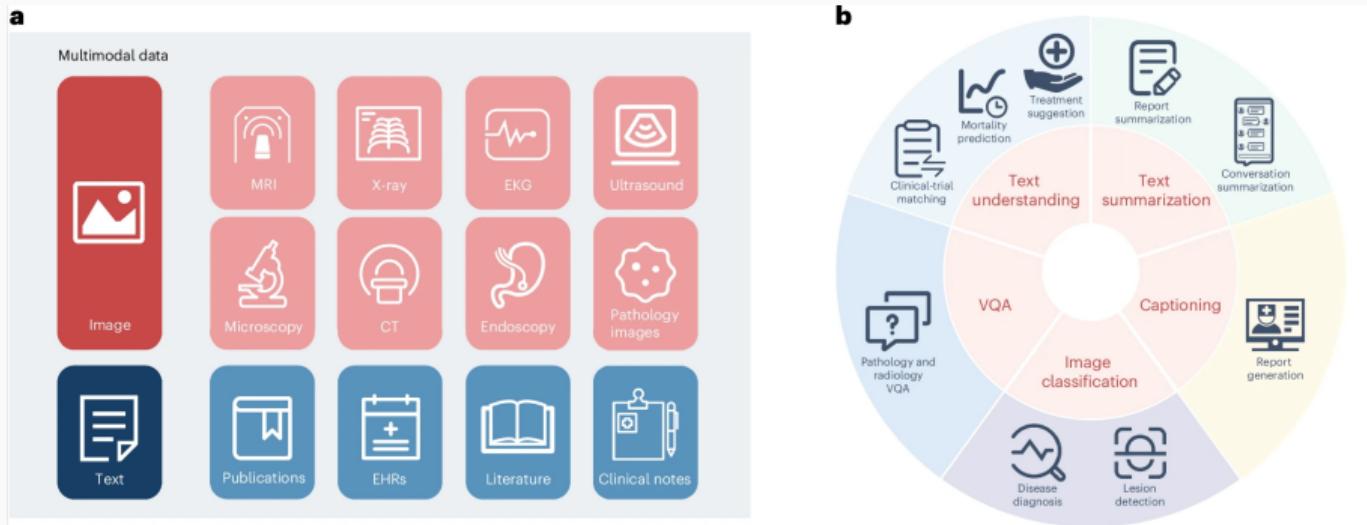
# Early fusion: blending information from the start



# Early fusion: blending information from the start



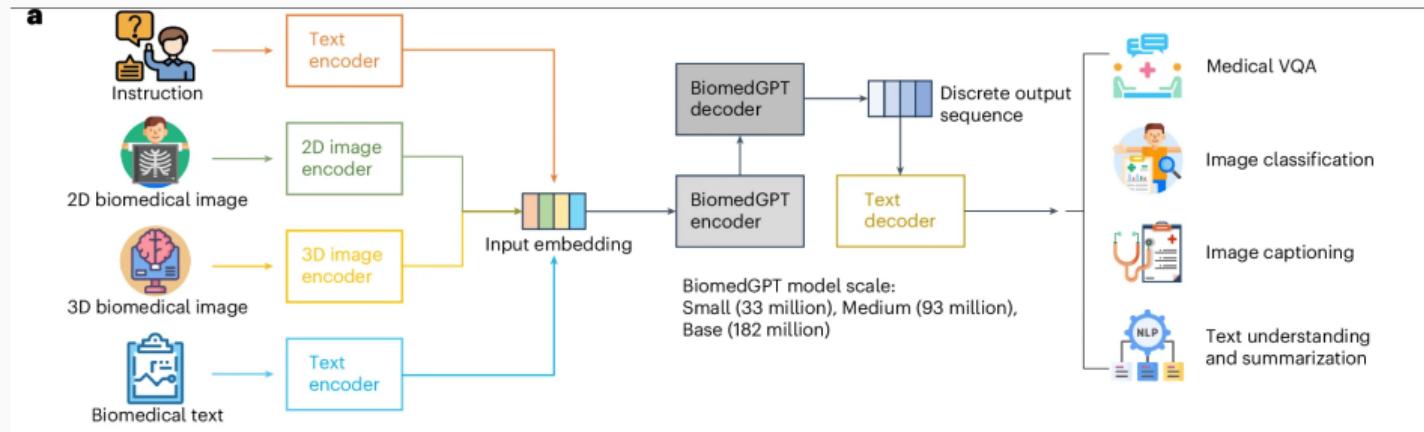
# Early fusion: blending information from the start



Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).



# Early fusion: blending information from the start



Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).



# Early fusion: blending information from the start



**c**

Pathology and radiology VQA



**Q:** What is seen at this stage, associated with regeneration and repair?  
**A:** Numerous reactive type II pneumocytes.

**Q:** Are bite cells like this one in the smear associated with regeneration and repair at this stage?  
**A:** No.

Report generation



**Q:** What are the findings based on the image?  
**A:** The nasogastric tube is in adequate position, and there is a resolution of the gastric distention. There is still mild bibasilar atelectasis. There are no pneumothorax or pleural effusion.

Disease diagnosis



**Q:** What disease does this image depict?  
**A:** Breast cancer.

Lesion detection



**Q:** What skin lesion does this image depict?  
**A:** Melanoma.



Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).



# Early fusion: blending information from the start



**g** Average zero-shot accuracy (%) across seven question types

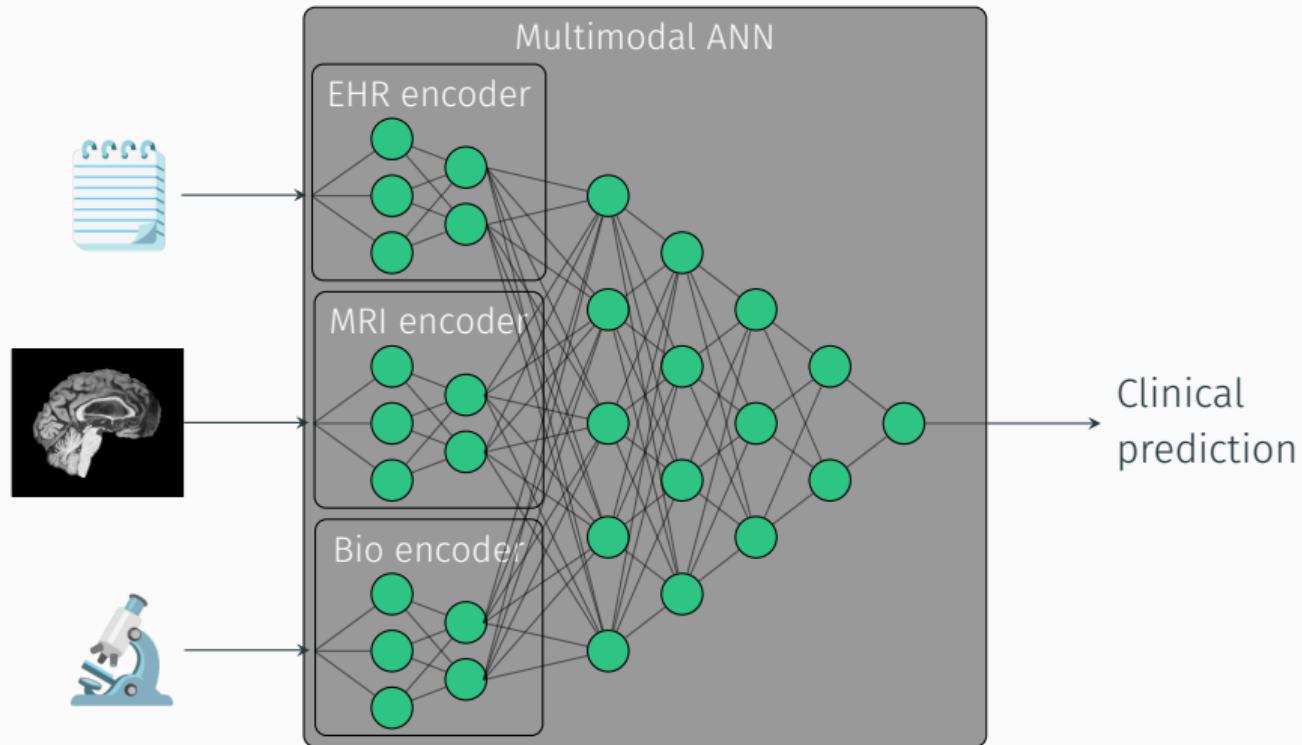
	GPT-4V	BiomedGPT-B	BiomedGPT-M	BiomedGPT-S	OFA-large	Instruct-BiomedGPT-B	Instruct-BiomedGPT-M	Instruct-BiomedGPT-S	LLaVA-med	OFA-huge
Disease diagnosis	50.9	43.5	35.4	45.2	32.8	53.9	52.1	32.6	42.0	34.5
Imaging technical details	73.3	41.0	19.9	19.5	20.9	68.8	68.0	58.1	67.6	20.4
Lesion and abnormality detection	48.6	37.2	38.7	41.2	45.5	49.5	52.9	45.9	40.6	40.6
Modality recognition	77.9	68.7	59.6	42.7	43.4	77.1	69.5	55.5	69.4	55.0
Size assessment	46.6	39.7	59.4	37.8	42.9	44.6	65.3	39.5	68.6	44.6
Spatial relationships	47.7	14.4	21.8	9.5	23.6	44.0	31.8	27.6	35.4	28.2
Structural identification	52.0	41.3	28.8	32.4	30.7	43.1	35.2	37.0	41.0	40.0



Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).



# Intermediate fusion: integrating insights along the way



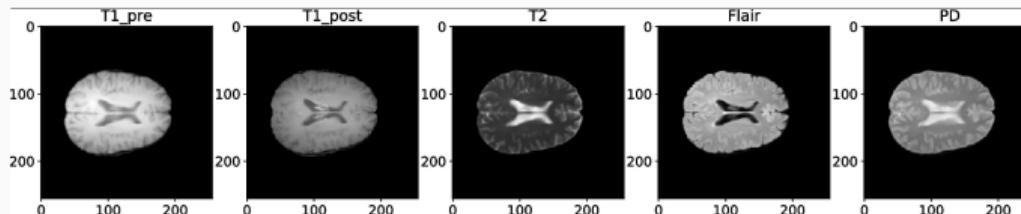
# Intermediate fusion: integrating insights along the way



## Structured EHR

LABORATORY TEST		VITAL SIGN		MEDICATION
Mean Corpuscular Hemoglobin	Carbon Dioxide	Albumin	Diastolic Blood Pressure	Baclofen
Red Cell Distribution Width	Basophils	Glucose Level	Systolic Blood Pressure	Gabapentin
Mean Corpuscular Hemoglobin Concentration	White Blood Cell Count	eGFR	Heart Rate	Copaxone
Mean Corpuscular Volume	Hematocrit	Albumin/Globulin Ratio	Weight	Gilenya

## Multimodal MRI



## Clinical notes

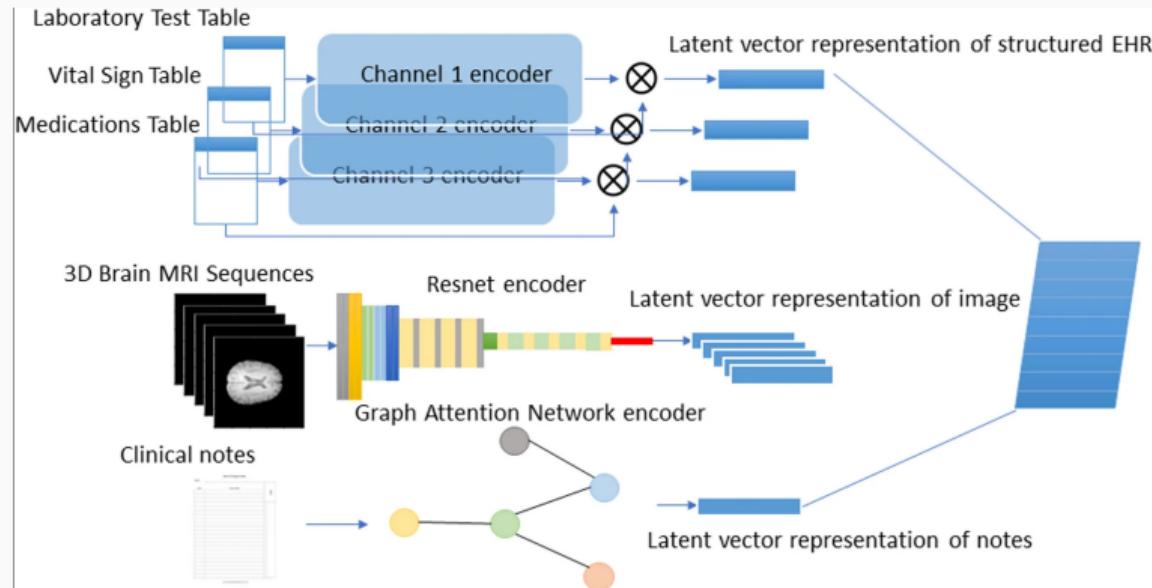
The patient's clinical notes are documented in unstructured free-text format and provide a comprehensive account of the patient's health status. These notes encompass a range of vital



Zhang, K., et al., Predicting multiple sclerosis severity with multimodal deep neural networks, *BMC Medical Informatics and Decision Making* (2023).



# Intermediate fusion: integrating insights along the way



Zhang, K., et al., Predicting multiple sclerosis severity with multimodal deep neural networks, *BMC Medical Informatics and Decision Making* (2023).



# Intermediate fusion: integrating insights along the way



	AUROC	AUPRC	Sensitivity	Specificity	Accuracy
MRI T1-pre	$0.6462 \pm 0.0352$	$0.2074 \pm 0.0145$	$0.5089 \pm 0.0397$	$0.7679 \pm 0.0209$	$0.6567 \pm 0.0300$
MRI T1-post	$0.6437 \pm 0.0389$	$0.2027 \pm 0.0180$	$0.5501 \pm 0.0390$	$0.6533 \pm 0.0252$	$0.6697 \pm 0.0199$
MRI T2	$0.7736 \pm 0.0268$	$0.2245 \pm 0.0198$	$0.6834 \pm 0.0223$	$0.7409 \pm 0.0398$	$0.7467 \pm 0.0390$
MRI FLAIR	$0.7945 \pm 0.2798$	$0.3306 \pm 0.0309$	<b><math>0.7689 \pm 0.0261</math></b>	$0.7423 \pm 0.0265$	$0.7423 \pm 0.0399$
MRI PD	$0.5430 \pm 0.0401$	$0.0998 \pm 0.0321$	<b><math>0.7536 \pm 0.0218</math></b>	$0.4862 \pm 0.0300$	$0.5046 \pm 0.0399$
Clinical Notes	$0.7048 \pm 0.0365$	$0.5201 \pm 0.0293$	$0.4632 \pm 0.0320$	<b><math>0.8956 \pm 0.0235</math></b>	$0.4958 \pm 0.0301$
Structured EHR	$0.6589 \pm 0.0193$	$0.3651 \pm 0.0265$	$0.7015 \pm 0.0263$	$0.6587 \pm 0.0366$	$0.6984 \pm 0.0265$
MRIs & Notes	<b><math>0.7988 \pm 0.0465</math></b>	<b><math>0.6321 \pm 0.0299</math></b>	$0.7024 \pm 0.0536$	<b><math>0.7792 \pm 0.0563</math></b>	<b><math>0.7963 \pm 0.0422</math></b>
MRIs & EHR	$0.7836 \pm 0.0531$	$0.4265 \pm 0.0323$	$0.6789 \pm 0.0411$	$0.6875 \pm 0.0333$	$0.6841 \pm 0.0523$
EHR & Notes	<b><math>0.8078 \pm 0.0232</math></b>	<b><math>0.7978 \pm 0.0453</math></b>	$0.7268 \pm 0.0435$	$0.7643 \pm 0.0255$	<b><math>0.8125 \pm 0.0353</math></b>
MS-BERT( [11] )	$0.6010 \pm 0.0222$	$0.2064 \pm 0.0356$	$0.3090 \pm 0.0265$	$0.7936 \pm 0.0512$	$0.7788 \pm 0.0398$
MRI & Notes & EHR	<b><math>0.8380 \pm 0.0438</math></b>	<b><math>0.7963 \pm 0.0520</math></b>	<b><math>0.7489 \pm 0.0502</math></b>	<b><math>0.7936 \pm 0.0488</math></b>	<b><math>0.7960 \pm 0.0312</math></b>



Zhang, K., et al., Predicting multiple sclerosis severity with multimodal deep neural networks, *BMC Medical Informatics and Decision Making* (2023).



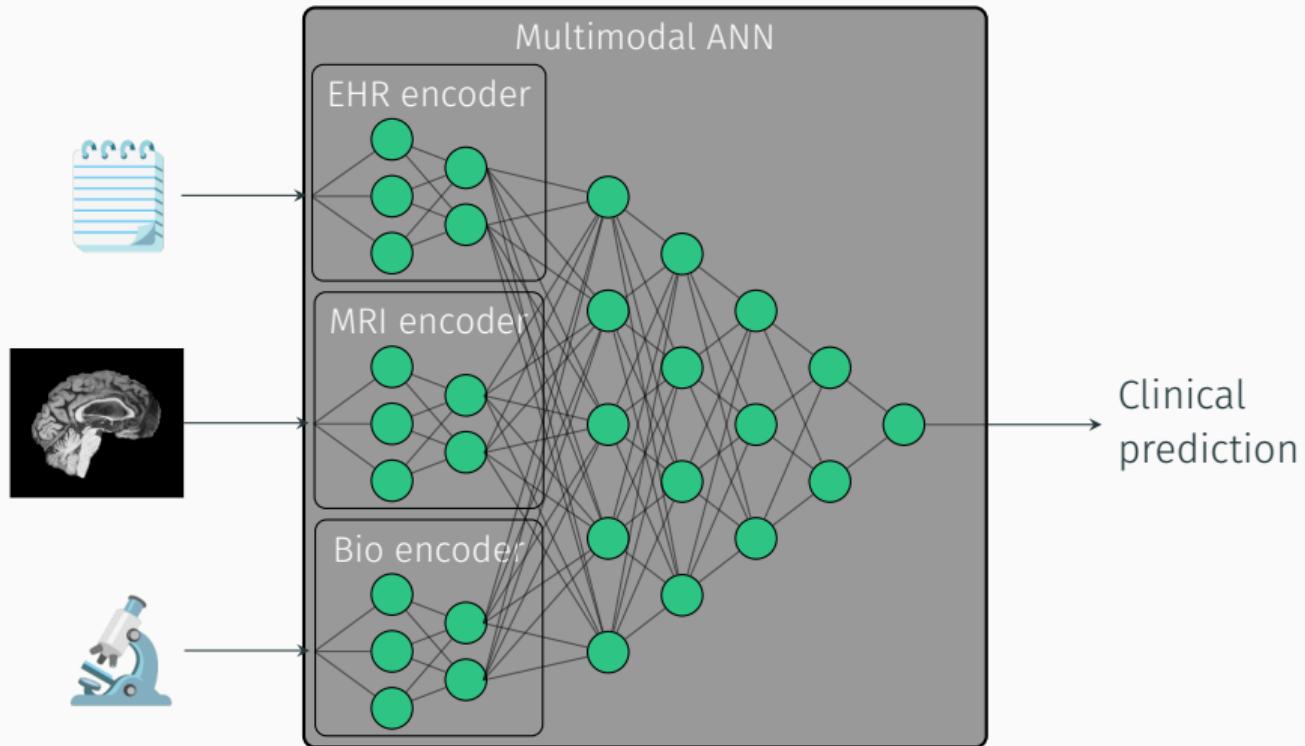
# The black-box problem of modern AI systems

---

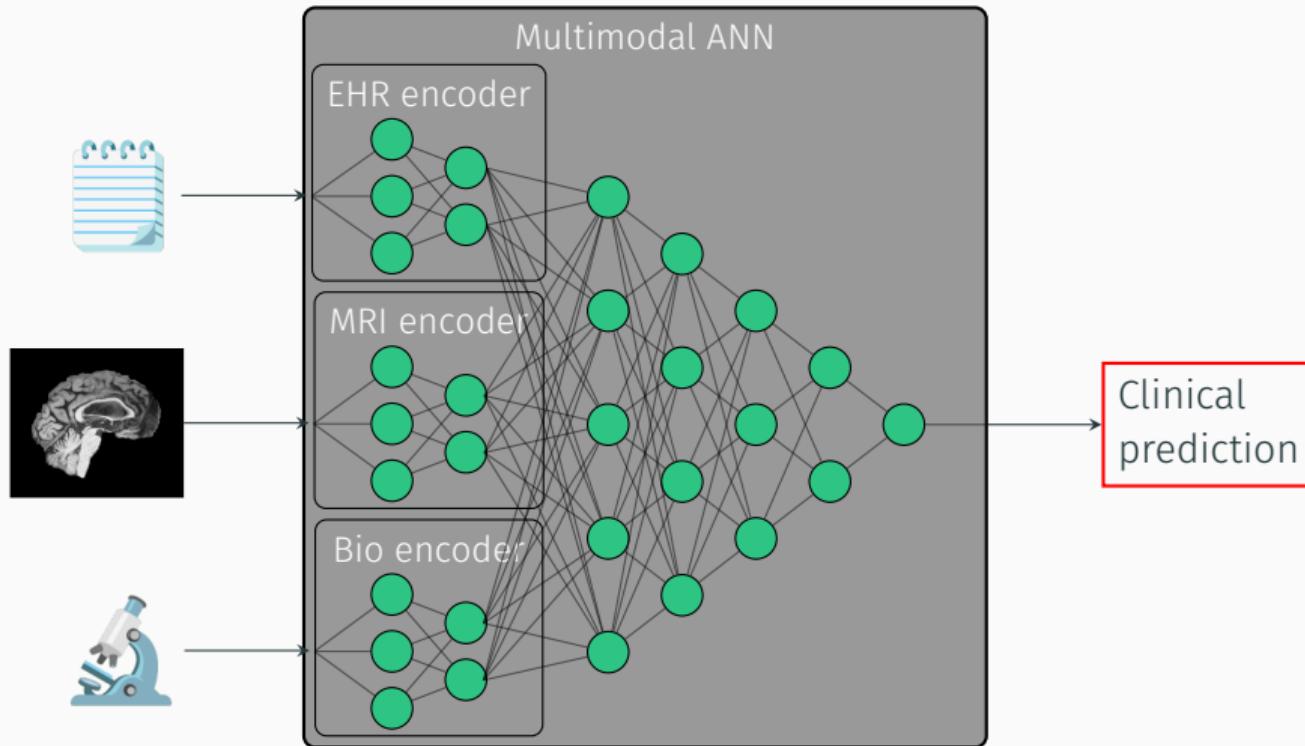


UNIVERSITETET  
I OSLO

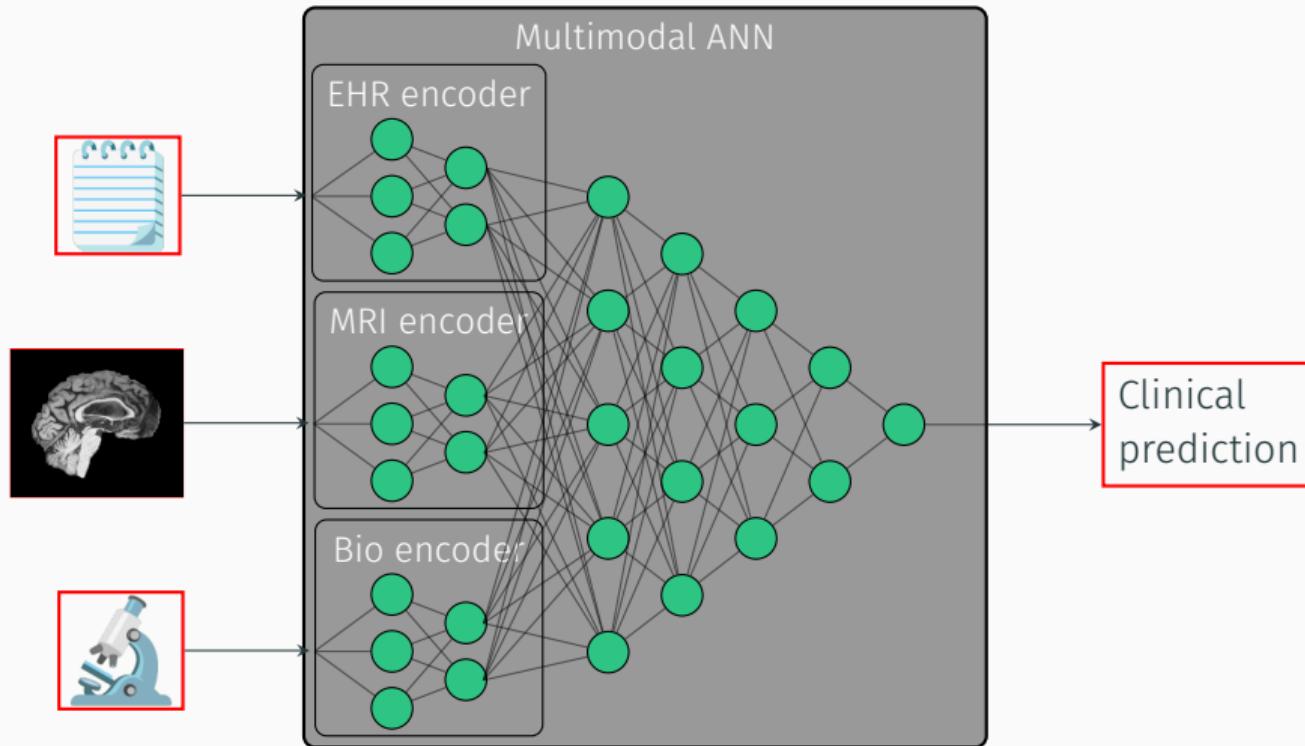
# The black-box problem of modern AI systems



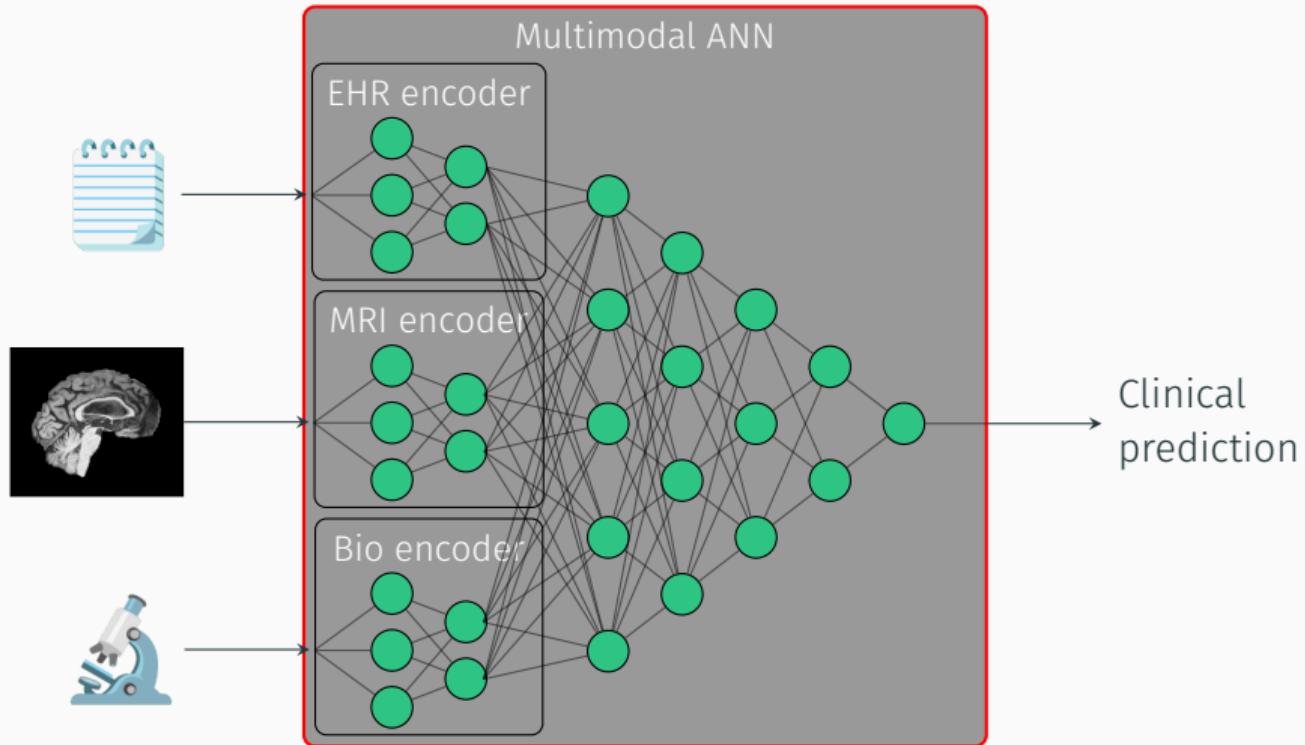
# The black-box problem of modern AI systems



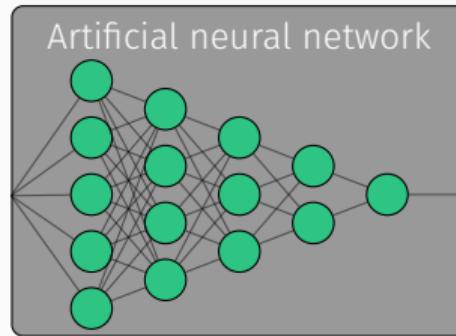
# The black-box problem of modern AI systems



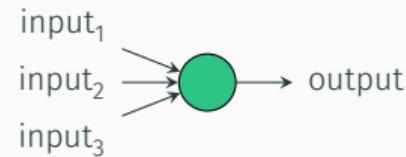
# The black-box problem of modern AI systems



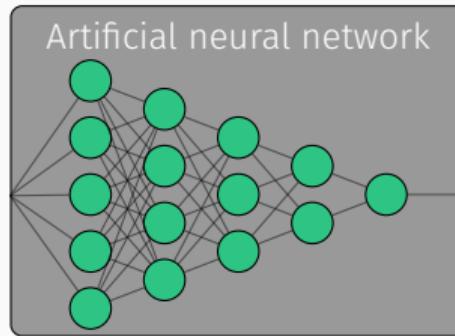
# The black-box problem of modern AI systems



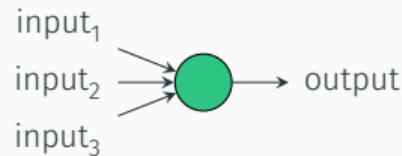
Artificial neuron



# The black-box problem of modern AI systems



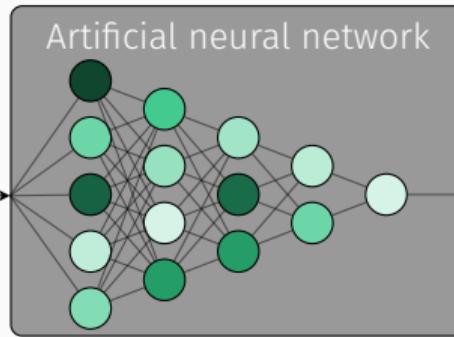
Artificial neuron



$$\text{output} = \max(0, \text{input}_1 * w_1 + \text{input}_2 * w_2 + \text{input}_3 * w_3)$$



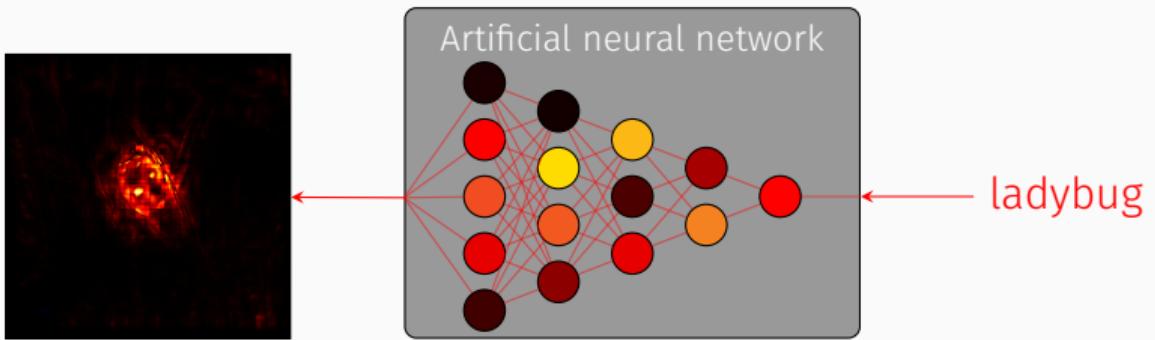
# The black-box problem of modern AI systems



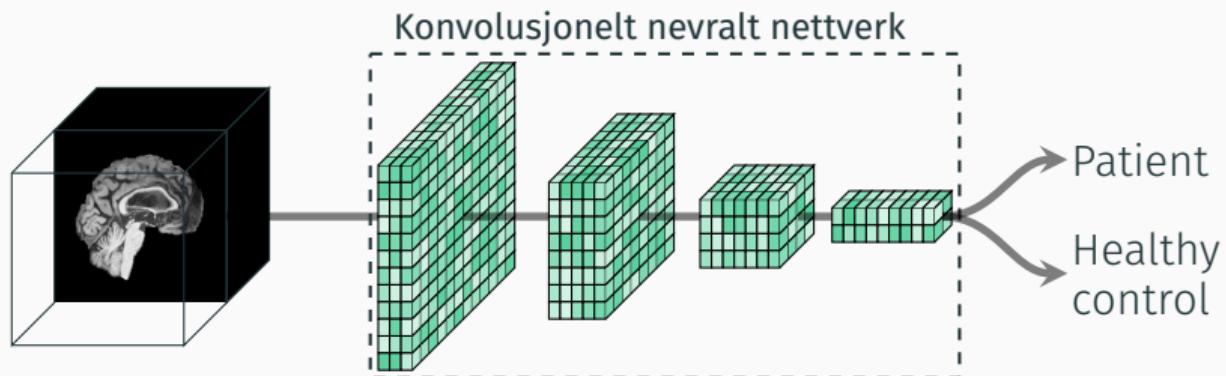
ladybug



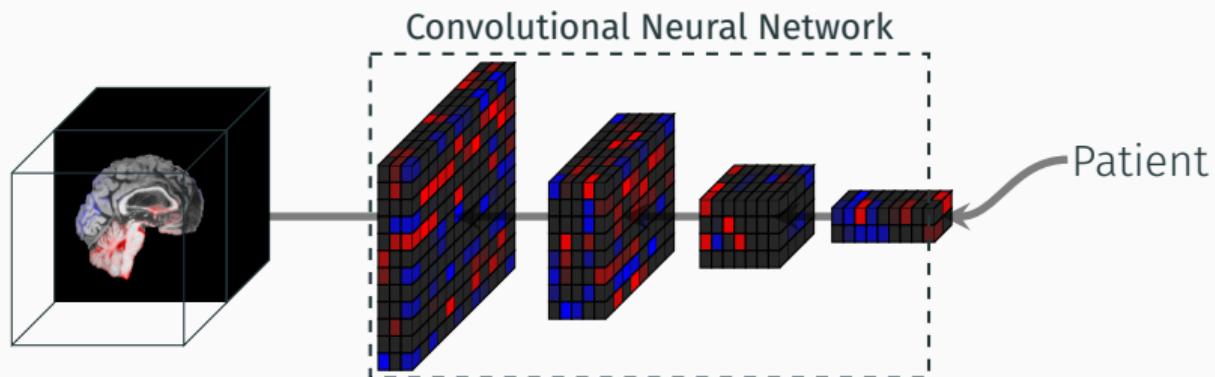
# The black-box problem of modern AI systems



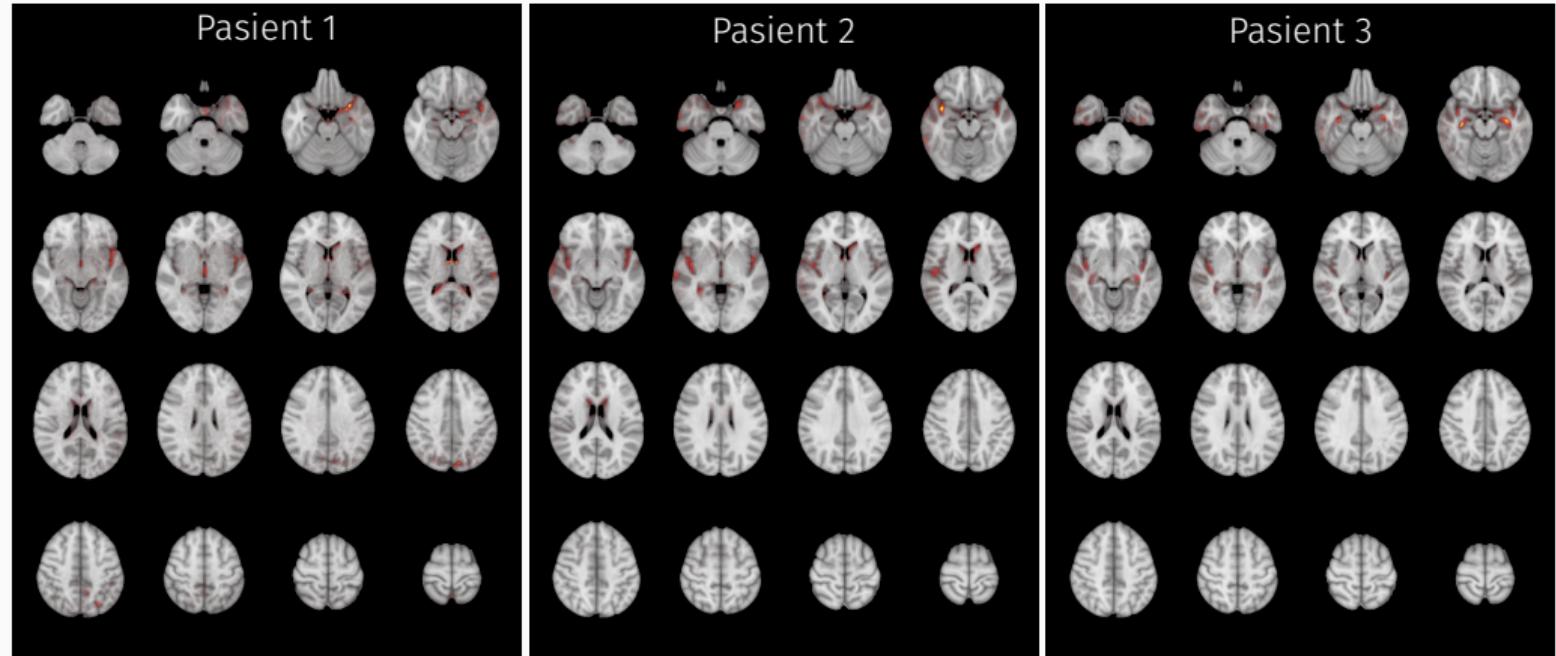
# The black-box problem of modern AI systems



# The black-box problem of modern AI systems

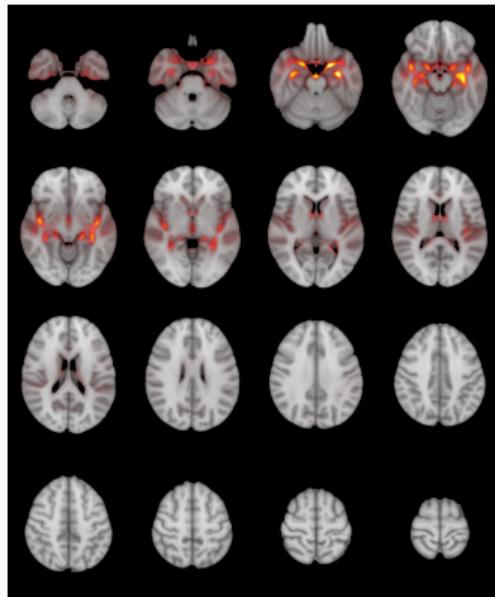


# Explainable AI and dementia



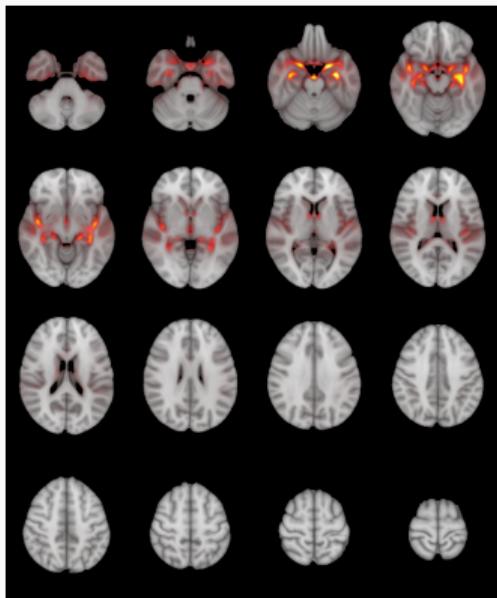
# Explainable AI and dementia

Forklarbar KI

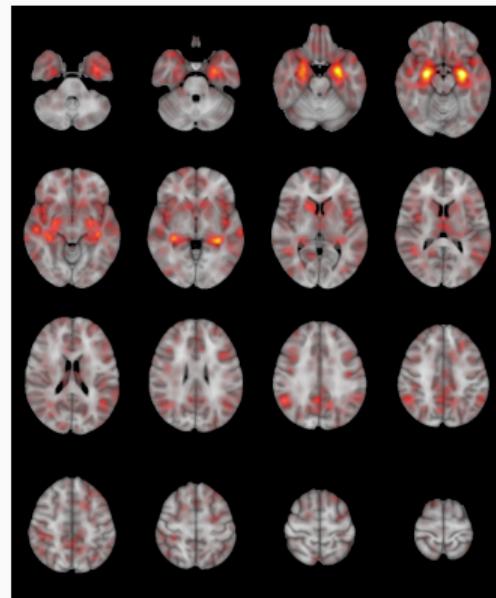


# Explainable AI and dementia

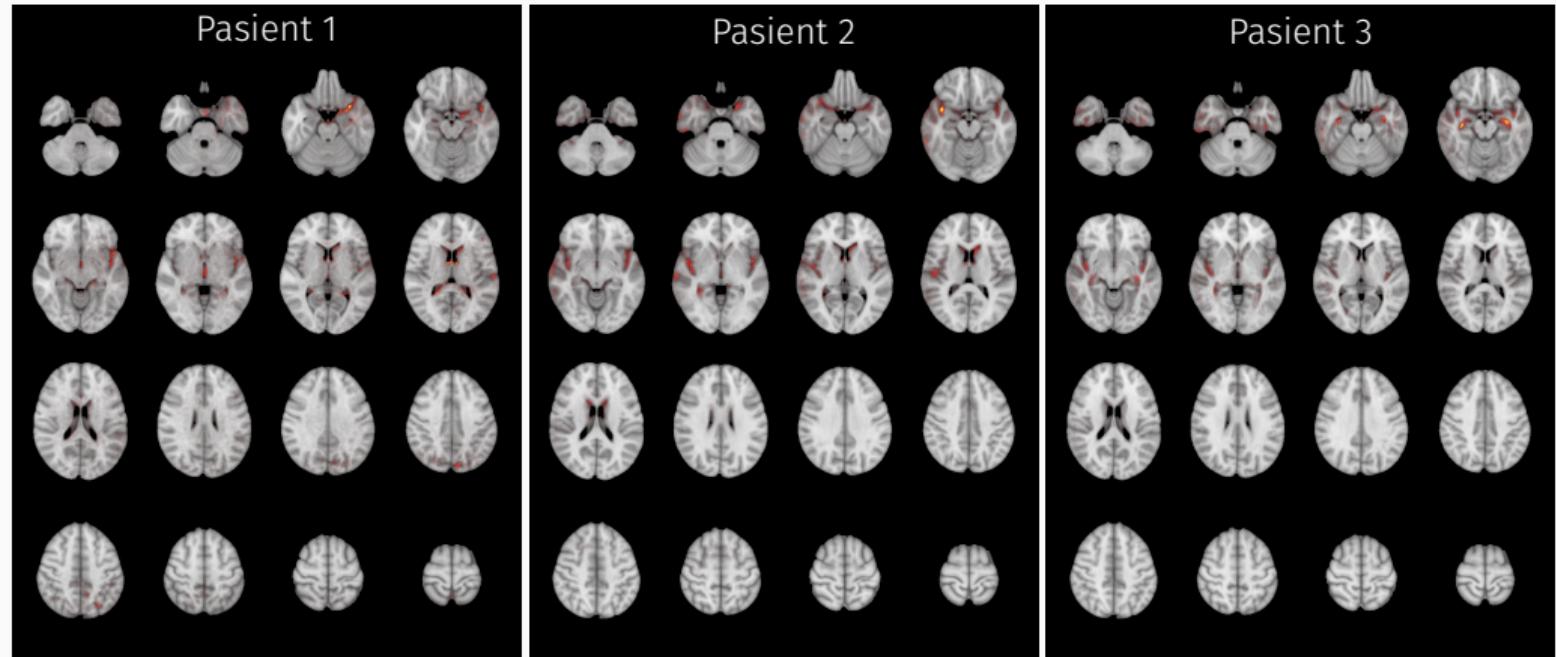
Forklarbar KI



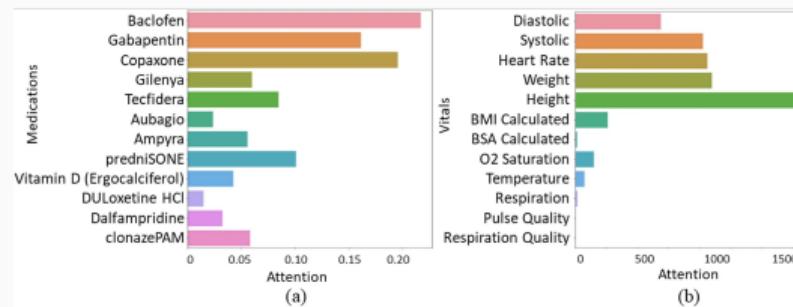
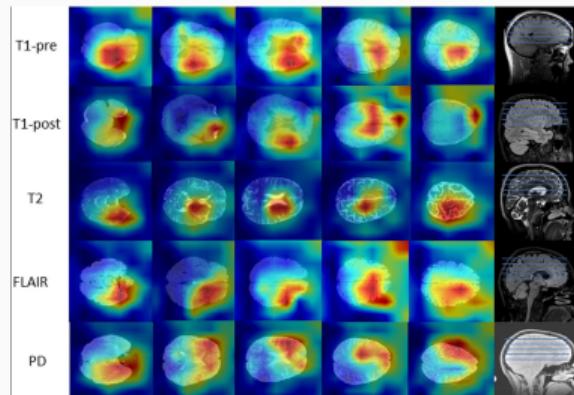
Mennesker



# Explainable AI and dementia



# Explainable AI and multimodality



# Explainable AI and multimodality



## Summary

- Deep learning is transforming many fields, enabling complex modelling of diverse, unstructured data.
- Multimodal AI requires decisions about how and when to combine information.
  - Late fusion: Information is merged after the most complex modelling
  - Early fusion: Information is merged before the most complex modelling
  - Intermediate fusion: Information is merged as a part of the most complex modelling
- Multimodal AI systems may enable clinical predictions with an accuracy surpassing what is currently possible, but explainability remains a challenge
  - Methods are emerging to alleviate these problems, at least partially



Thank you for your attention!  
estenhl@ui.no



UNIVERSITETET  
I OSLO

