Integrating complex multimodal health data for clinical prediction with artificial intelligence

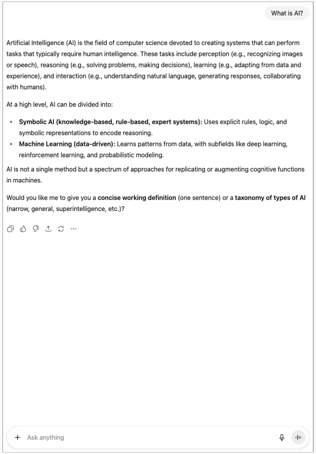
Esten H. Leonardsen

Post-doktor ved Psykologisk institutt, Universitet i Oslo

Chief Scientific Officer, baba.vision

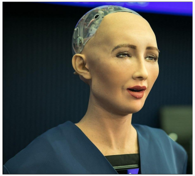
UNIVERSITETET I OSLO

What is artificial intelligence?

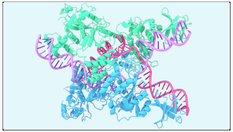


ChatGPT

Spot



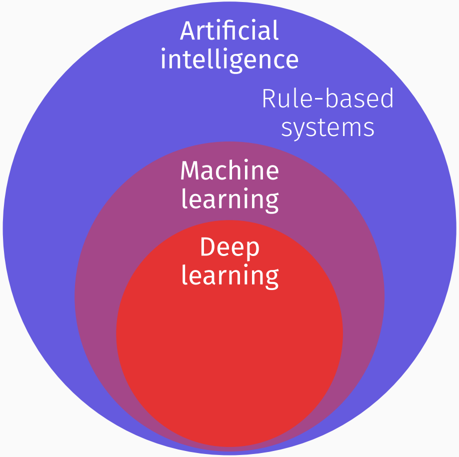
Sophia



AlphaFold

AlphaZero

What is artificial intelligence?

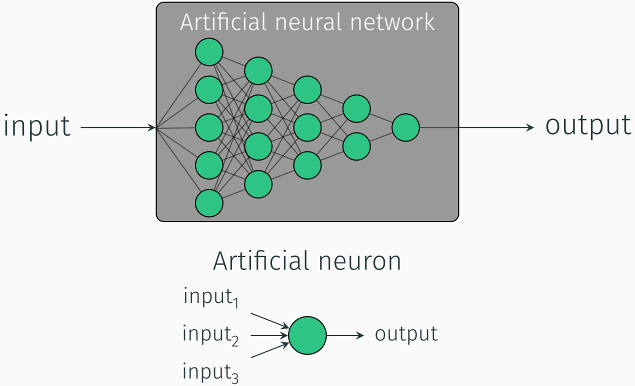


Artificial intelligence
The field of study producing technology that solves tasks requiring some form of intelligence

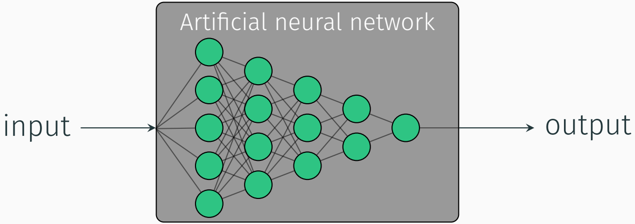
Machine learning
A set of techniques to solve problems by allowing machines to find patterns in training data

Deep learning
Machine learning approaches that rely on artificial neural networks

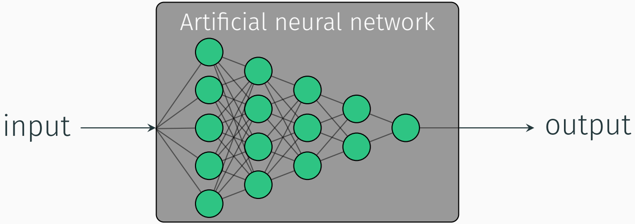
Why do we use artificial neural networks?

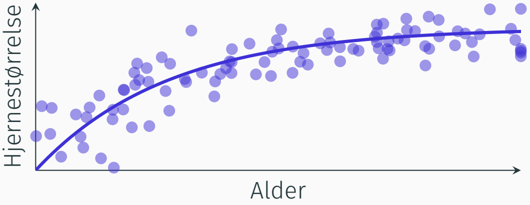


Why do we use artificial neural networks?

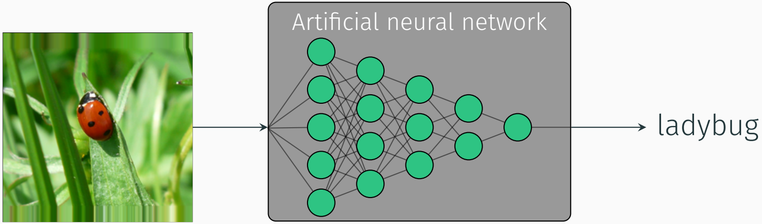


Why do we use artificial neural networks?

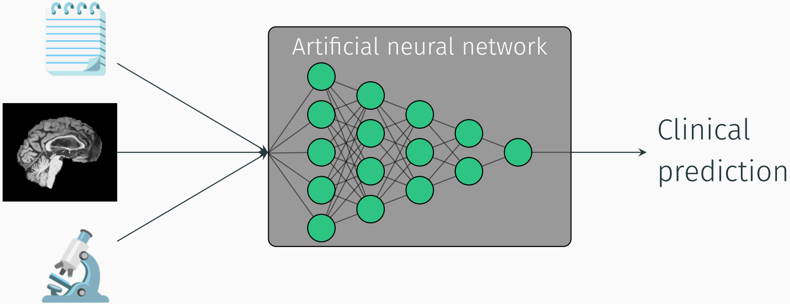




Why do we use artificial neural networks?



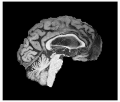
Why do we use artificial neural networks?



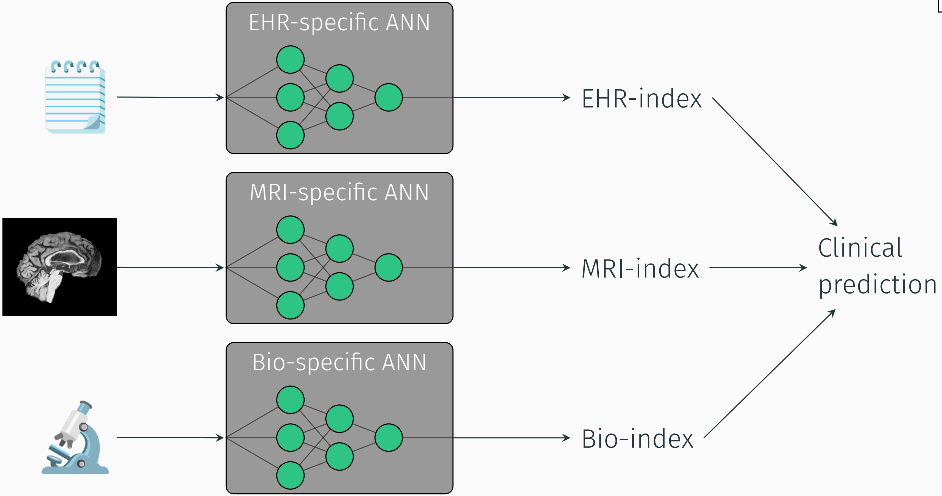
Integrating multimodal health data for clinical predictions

UNIVERSITETET
I OSLO

Late fusion: independent insights, combined decisions



Late fusion: independent insights, combined decisions



Late fusion: independent insights, combined decisions

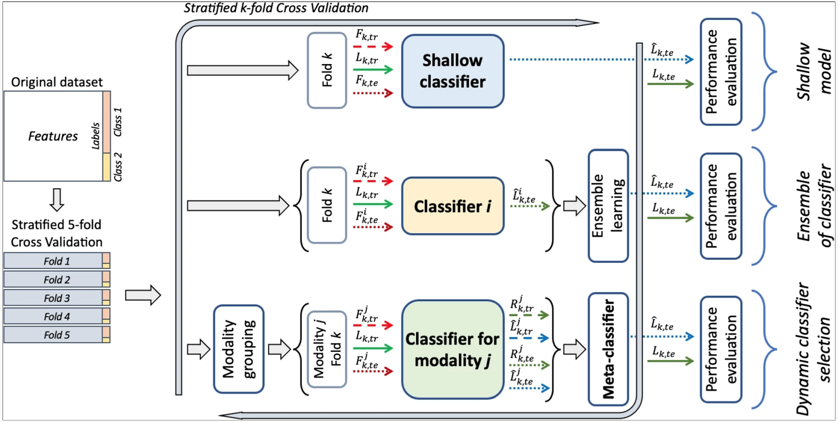
- Modality 1: Anamnestic data, which includes Age at hospitalization, Sex, Cardiovascular disorders, Neurological disorders, Oncological disorders, Diabetes, Obesity, Dyspnea and Smoking attitude;

- 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 (SpO_2) at hospitalization, Oxygen Saturation (O_2) at hospitalization, Days of ventilation, Continuous positive airway pressure (CPAP) ventilation and Non-invasive ventilation (NIV) ventilation.

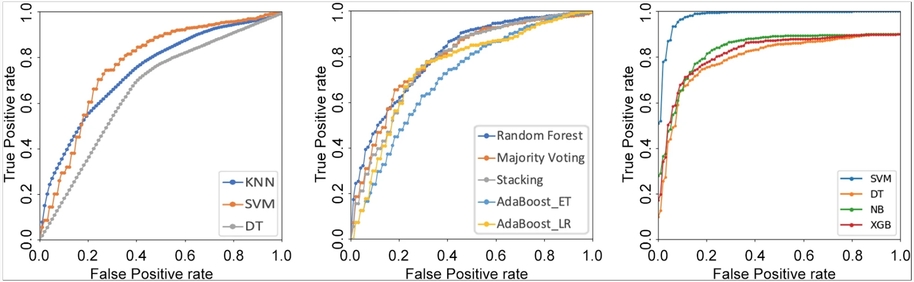
Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).

Late fusion: independent insights, combined decisions



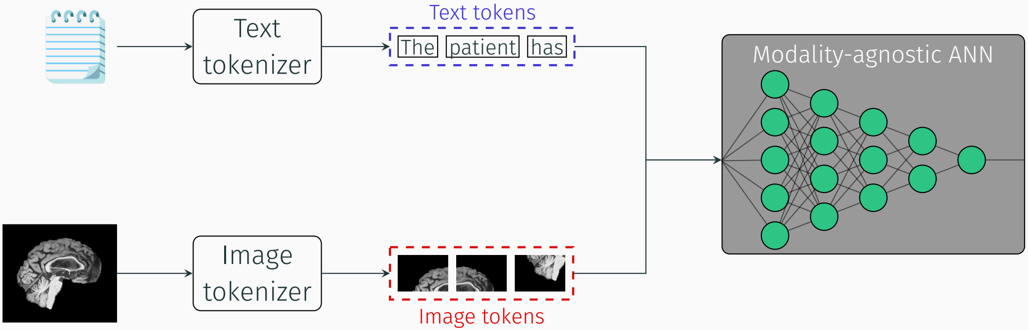
Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).

Late fusion: independent insights, combined decisions

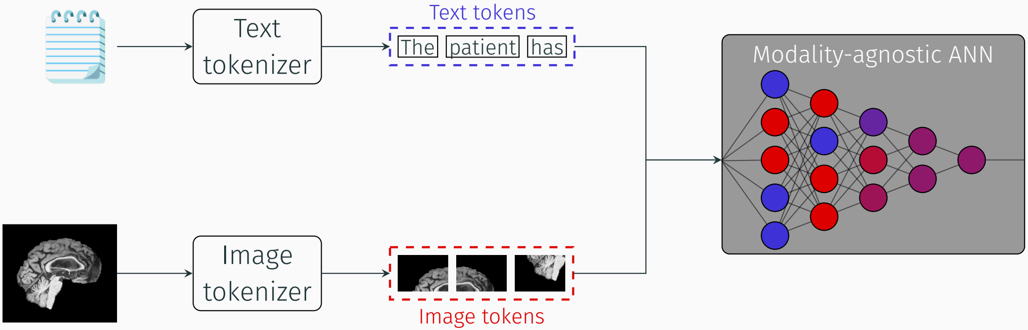


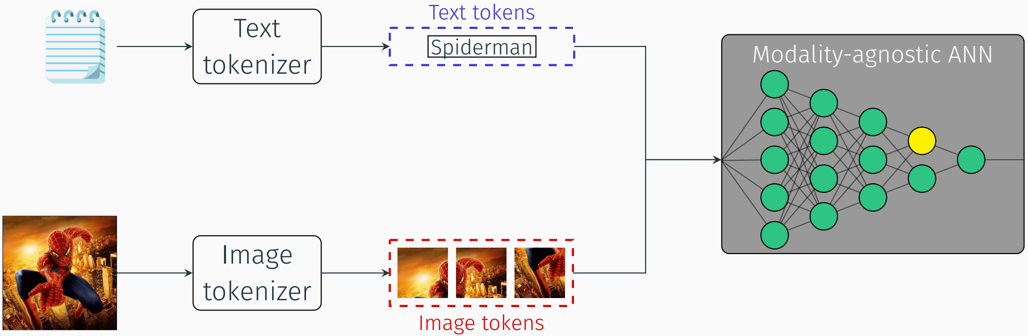
Cordelli, E., et al., Machine learning predicts pulmonary Long Covid sequelae using clinical data, *BMC Medical Informatics and Decision Making* (2024).

Early fusion: blending information from the start

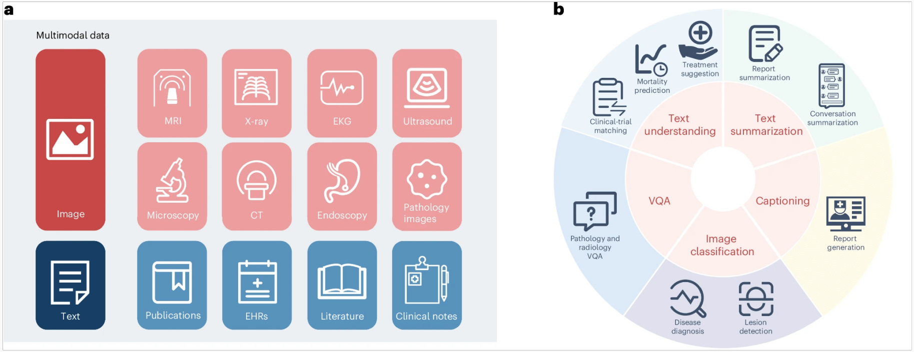


Early fusion: blending information from the start



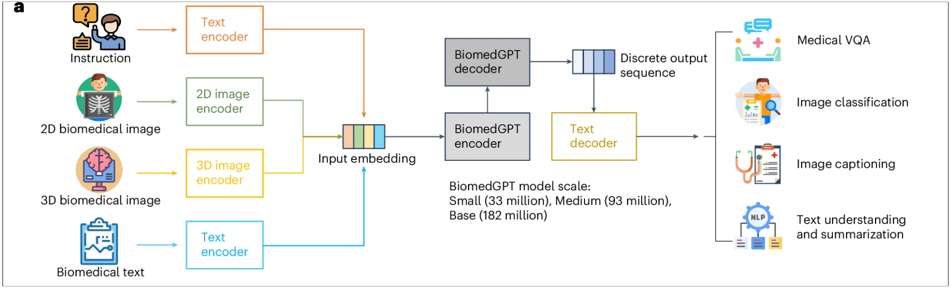

Text
tokenizer
Text tokens
Spiderman
Image
tokenizer
Image tokens
Modality-agnostic ANN

Early fusion: blending information from the start



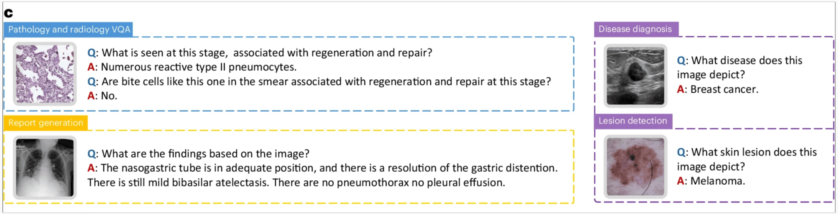
Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).

Early fusion: blending information from the start

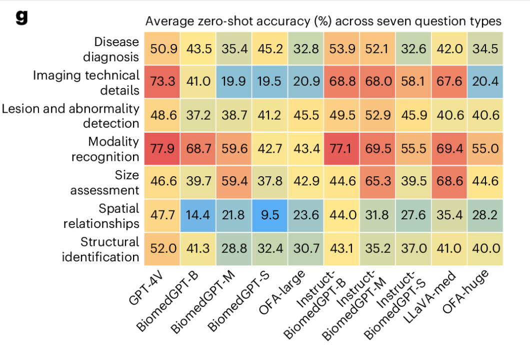


Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).

Early fusion: blending information from the start

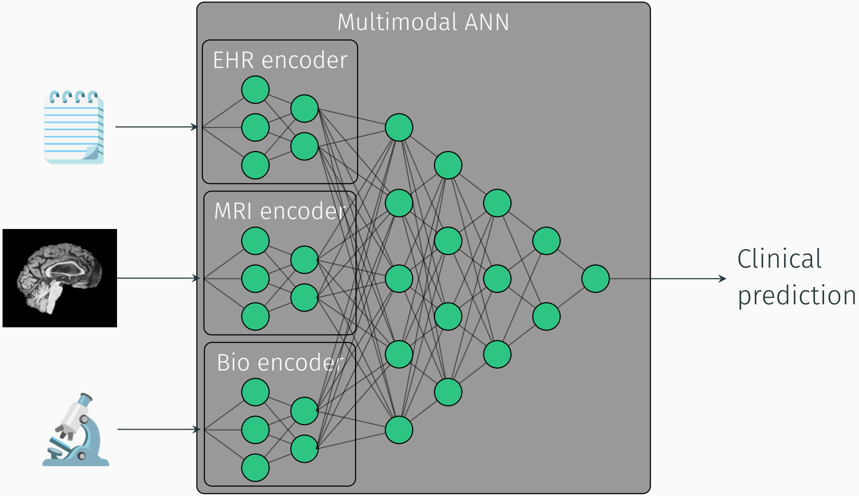


Zhang, K., et al., A generalist vision–language foundation model for diverse biomedical tasks, *Nature Medicine* (2024).



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

Intermediate fusion: integrating insights along the way

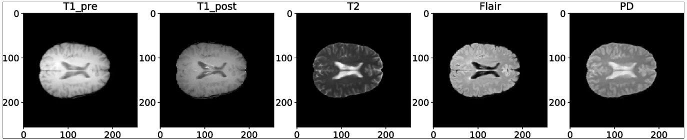


Intermediate fusion: integrating insights along the way

Structured EHR

LABORATORY TEST			VITAL SIGN	MEDICATION
Mean Corpuscular Hemoglobi	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

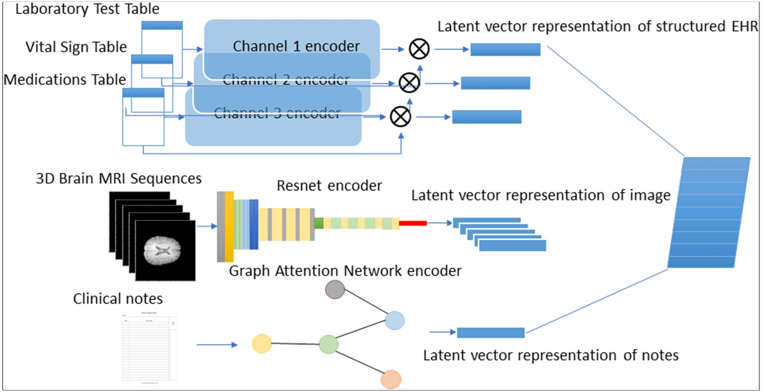


Clinical notes

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).

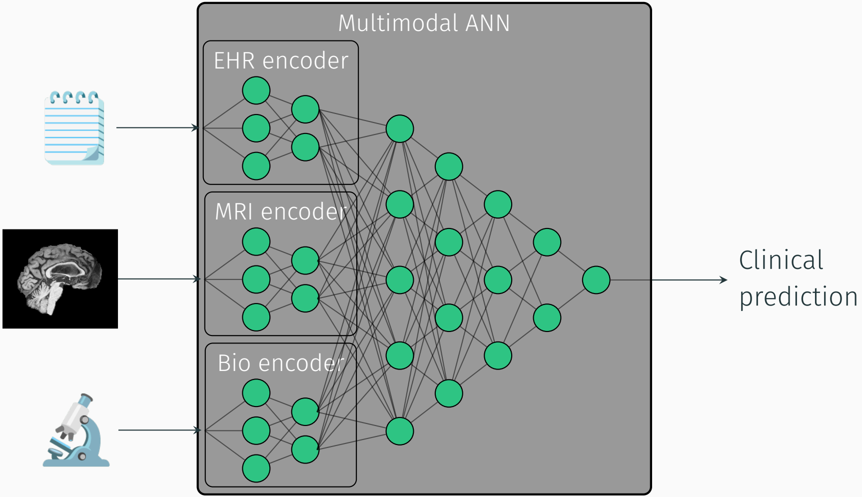
Intermediate fusion: integrating insights along the way

	AUROC	AUPRC	Sensitivity	Specificity	Accuracy
MRI T1-pre	0.6462 ± 0.0352	0.2074 ± 0.0145	0.5089 ± 0.0397	0.7679 ± 0.0209	0.6567 ± 0.0300
MRI T1-post	0.6437 ± 0.0389	0.2027 ± 0.0180	0.5501 ± 0.0390	0.6536 ± 0.0252	0.6697 ± 0.0199
MRI T2	0.7736 ± 0.0268	0.2245 ± 0.0198	0.6834 ± 0.0223	0.7409 ± 0.0398	0.7467 ± 0.0390
MRI FLAIR	0.7945 ± 0.2798	0.3306 ± 0.0309	**0.7689 ± 0.0261**	0.7423 ± 0.0265	0.7423 ± 0.0399
MRI PD	0.5430 ± 0.0401	0.0998 ± 0.0321	**0.7536 ± 0.0218**	0.4862 ± 0.0300	0.5046 ± 0.0399
Clinical Notes	0.7048 ± 0.0365	0.5201 ± 0.0293	0.4632 ± 0.0320	**0.8956 ± 0.0235**	0.4958 ± 0.0301
Structured EHR	0.6589 ± 0.0193	0.3651 ± 0.0265	0.7015 ± 0.0263	0.6587 ± 0.0366	0.6984 ± 0.0265
MRIs & Notes	**0.7988 ± 0.0465**	**0.6321 ± 0.0299**	0.7024 ± 0.0536	**0.7792 ± 0.0563**	**0.7963 ± 0.0422**
MRIs & EHR	0.7836 ± 0.0531	0.4265 ± 0.0323	0.6789 ± 0.0411	0.6875 ± 0.0333	0.6841 ± 0.0523
EHR & Notes	**0.8078 ± 0.0232**	**0.7978 ± 0.0453**	0.7268 ± 0.0435	0.7643 ± 0.0255	**0.8125 ± 0.0353**
MS-BERT([11])	0.6010 ± 0.0222	0.2064 ± 0.0356	0.3090 ± 0.0265	0.7936 ± 0.0512	0.7788 ± 0.0398
MRI & Notes & EHR	**0.8380 ± 0.0438**	**0.7963 ± 0.0520**	**0.7489 ± 0.0502**	**0.7936 ± 0.0488**	**0.7960 ± 0.0312**

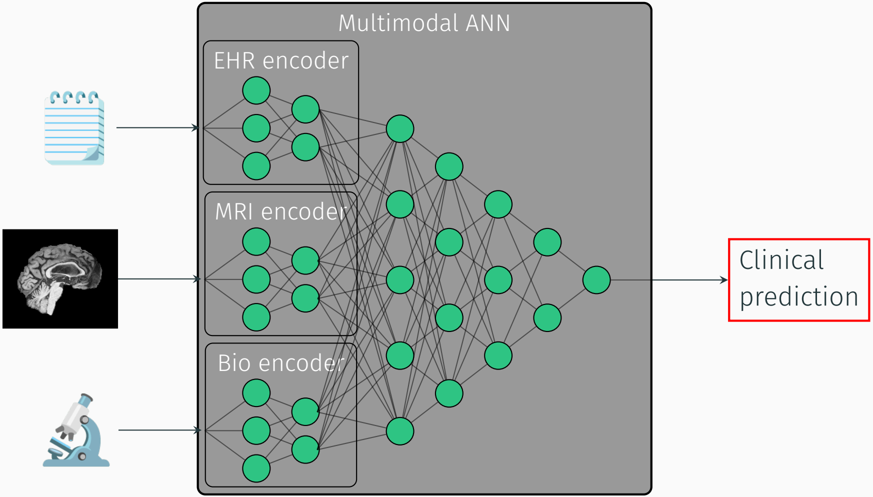
Zhang, K., et al., Predicting multiple sclerosis severity with multimodal deep neural networks, *BMC Medical Informatics and Decision Making* (2023).

The black-box problem of modern AI systems

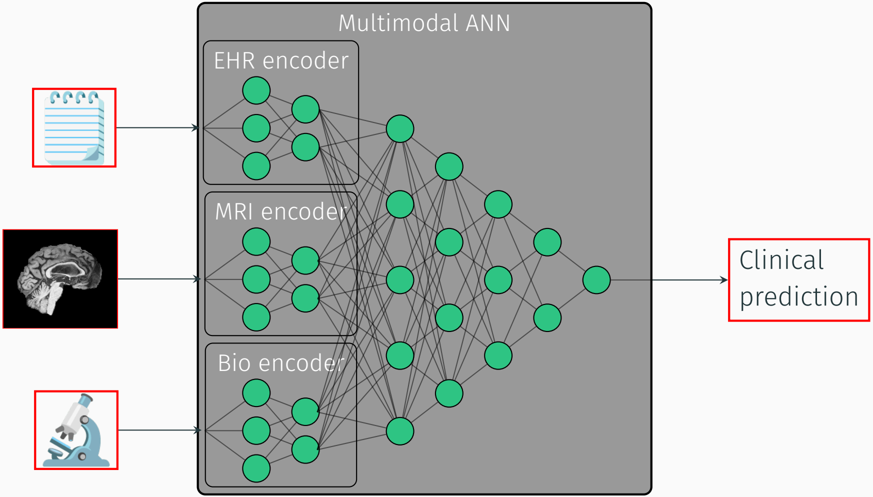
The black-box problem of modern AI systems



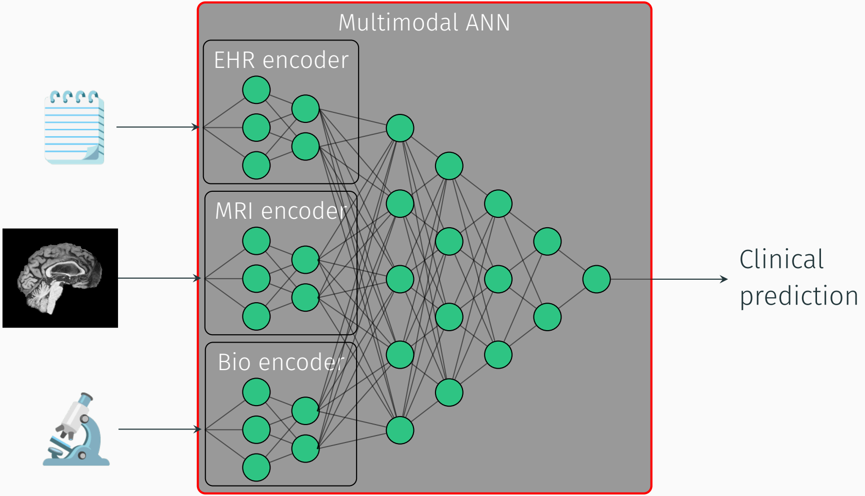
The black-box problem of modern AI systems



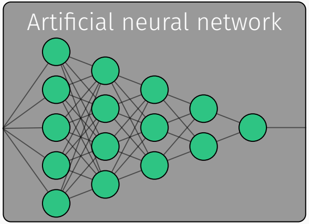
The black-box problem of modern AI systems

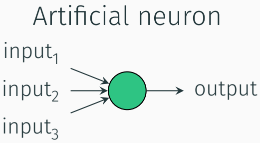


The black-box problem of modern AI systems

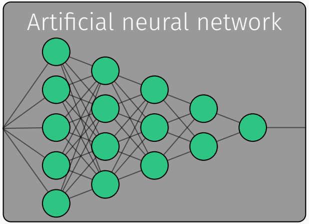


The black-box problem of modern AI systems

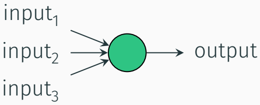




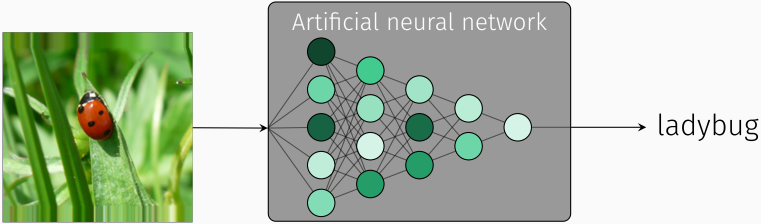
The black-box problem of modern AI systems



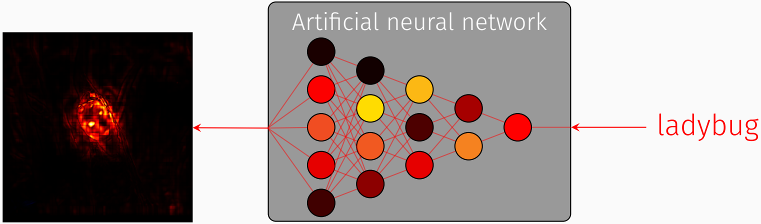
Artificial neuron



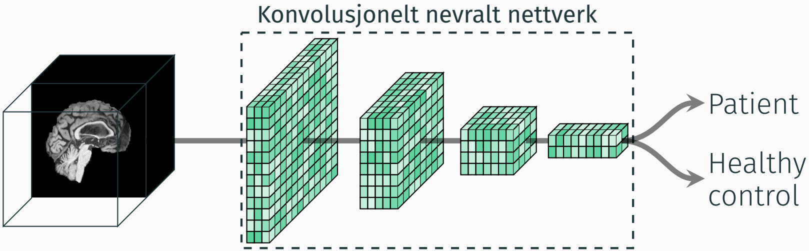
$$\text{output} = \max(0, \text{input}_1 * w_1 + \text{input}_2 * w_2 + \text{input}_3 * w_3)$$


Artificial neural network
ladybug

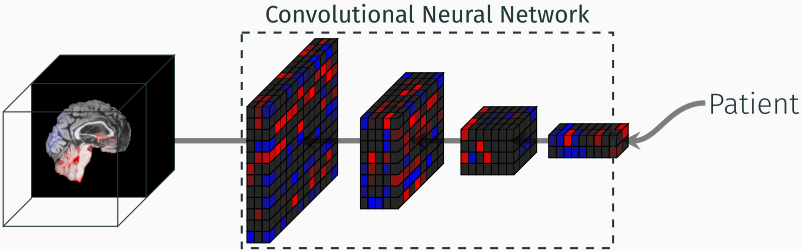
The black-box problem of modern AI systems



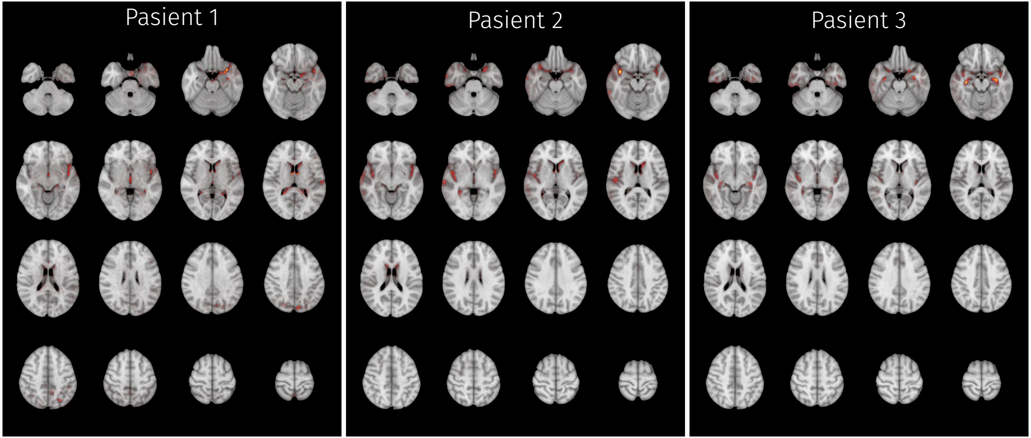
The black-box problem of modern AI systems



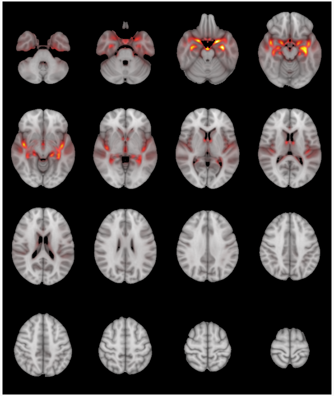
The black-box problem of modern AI systems



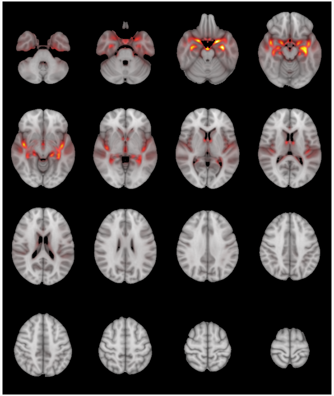
Explainable AI and dementia



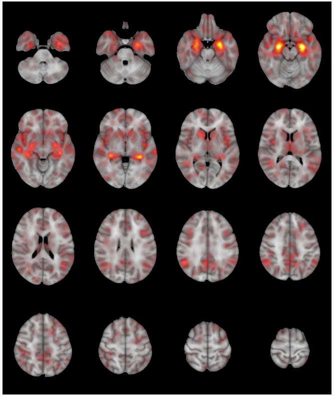
Forklarbar KI



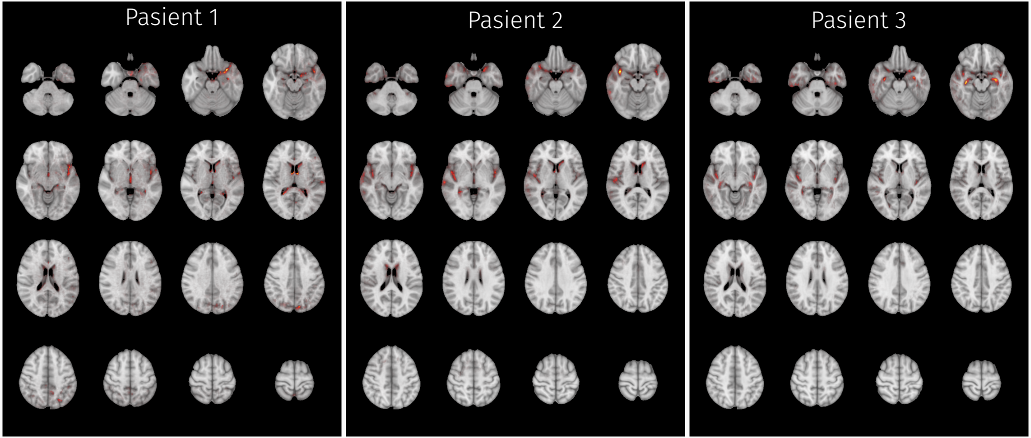
Forklarbar KI



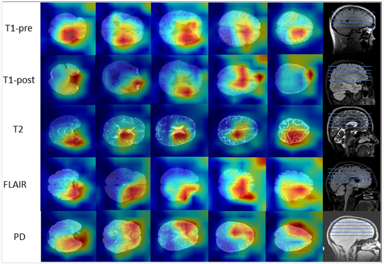
Mennesker

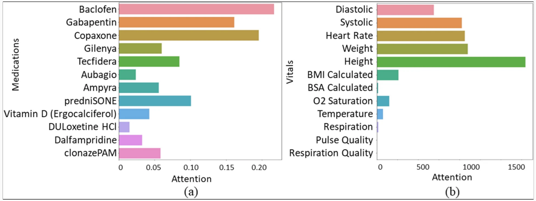


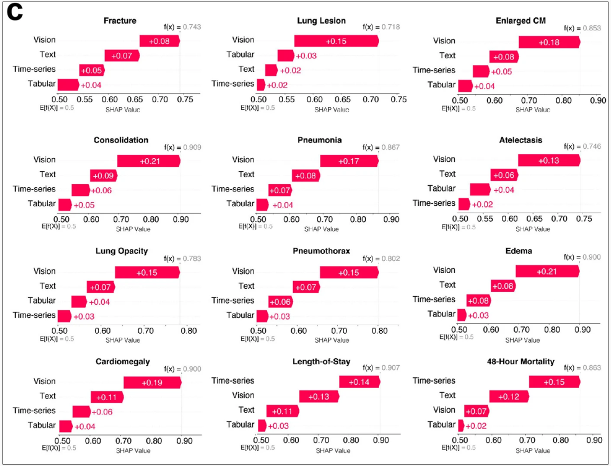
Explainable AI and dementia



Explainable AI and multimodality






c
Fracture
f(x) = 0.743
Vision +0.08
Text +0.07
Time-series +0.05
Tabular +0.04
0.50 0.55 0.60 0.65 0.70 0.75
E[f(X)] = 0.5
SHAP Value
Lung Lesion
f(x) = 0.718
Vision +0.15
Tabular +0.03
Text +0.02
Time-series +0.02
0.50 0.55 0.60 0.65 0.70 0.75
E[f(X)] = 0.5
SHAP Value
Enlarged CM
f(x) = 0.853
Vision +0.18
Text +0.08
Time-series +0.05
Tabular +0.04
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
Consolidation
f(x) = 0.909
Vision +0.21
Text +0.09
Time-series +0.06
Tabular +0.05
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
Pneumonia
f(x) = 0.867
Vision +0.17
Text +0.08
Time-series +0.07
Tabular +0.04
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
Atelectasis
f(x) = 0.746
Vision +0.13
Text +0.06
Tabular +0.04
Time-series +0.02
0.50 0.55 0.60 0.65 0.70 0.75
E[f(X)] = 0.5
SHAP Value
Lung Opacity
f(x) = 0.783
Vision +0.15
Text +0.07
Tabular +0.04
Time-series +0.03
0.50 0.60 0.70 0.80
E[f(X)] = 0.5
SHAP Value
Pneumothorax
f(x) = 0.802
Vision +0.15
Text +0.07
Time-series +0.06
Tabular +0.03
0.50 0.60 0.70 0.80
E[f(X)] = 0.5
SHAP Value
Edema
f(x) = 0.900
Vision +0.21
Text +0.08
Time-series +0.08
Tabular +0.03
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
Cardiomegaly
f(x) = 0.900
Vision +0.19
Text +0.11
Time-series +0.06
Tabular +0.04
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
Length-of-Stay
f(x) = 0.907
Time-series +0.14
Vision +0.13
Text +0.11
Tabular +0.03
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value
48-Hour Mortality
f(x) = 0.863
Time-series +0.15
Text +0.12
Vision +0.07
Tabular +0.02
0.50 0.60 0.70 0.80 0.90
E[f(X)] = 0.5
SHAP Value

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@uio.no