

KUNSTIG INTELLIGENS OG BIVIRKNINGER

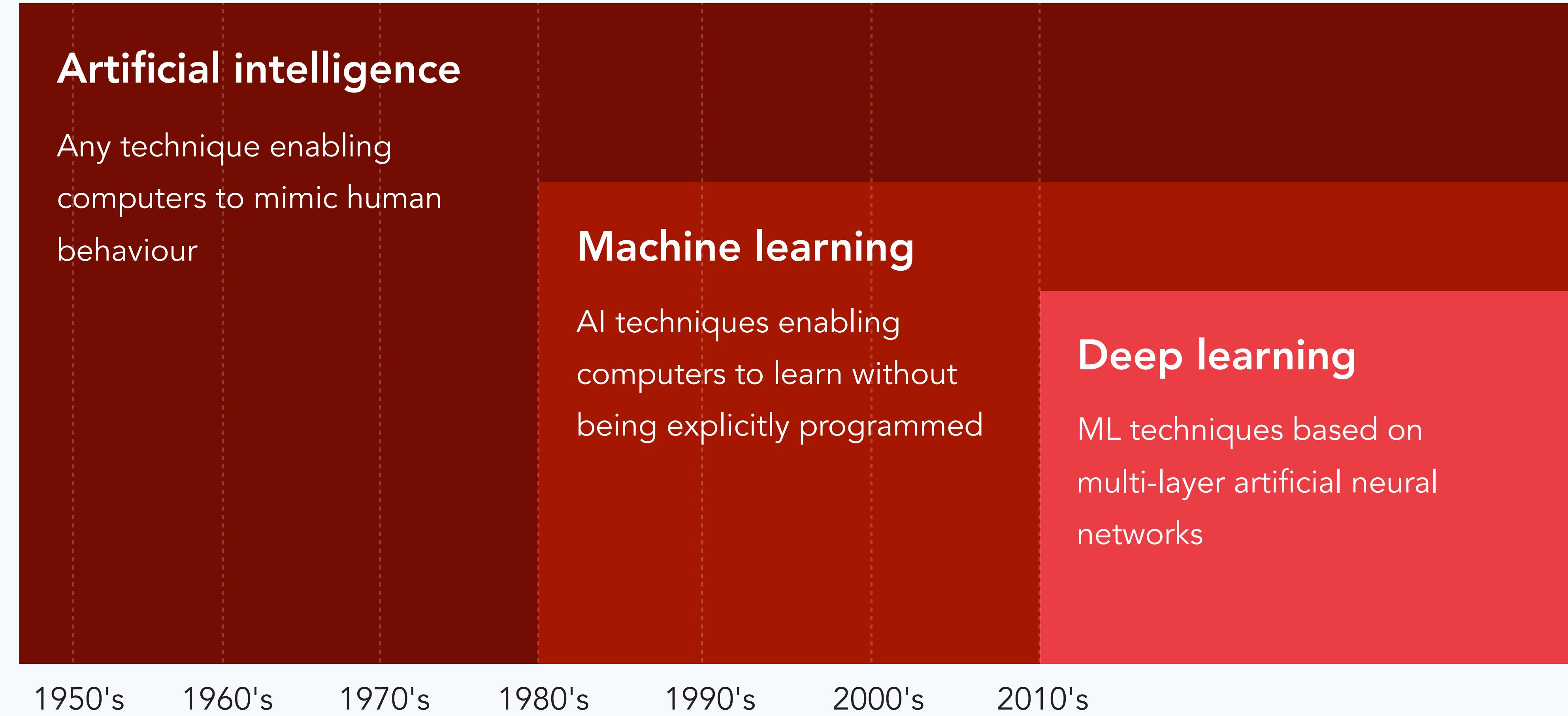
– ELLER HVORDAN MAN LÆSER
3 MILLIONER JOURNALER PÅ EN UGE

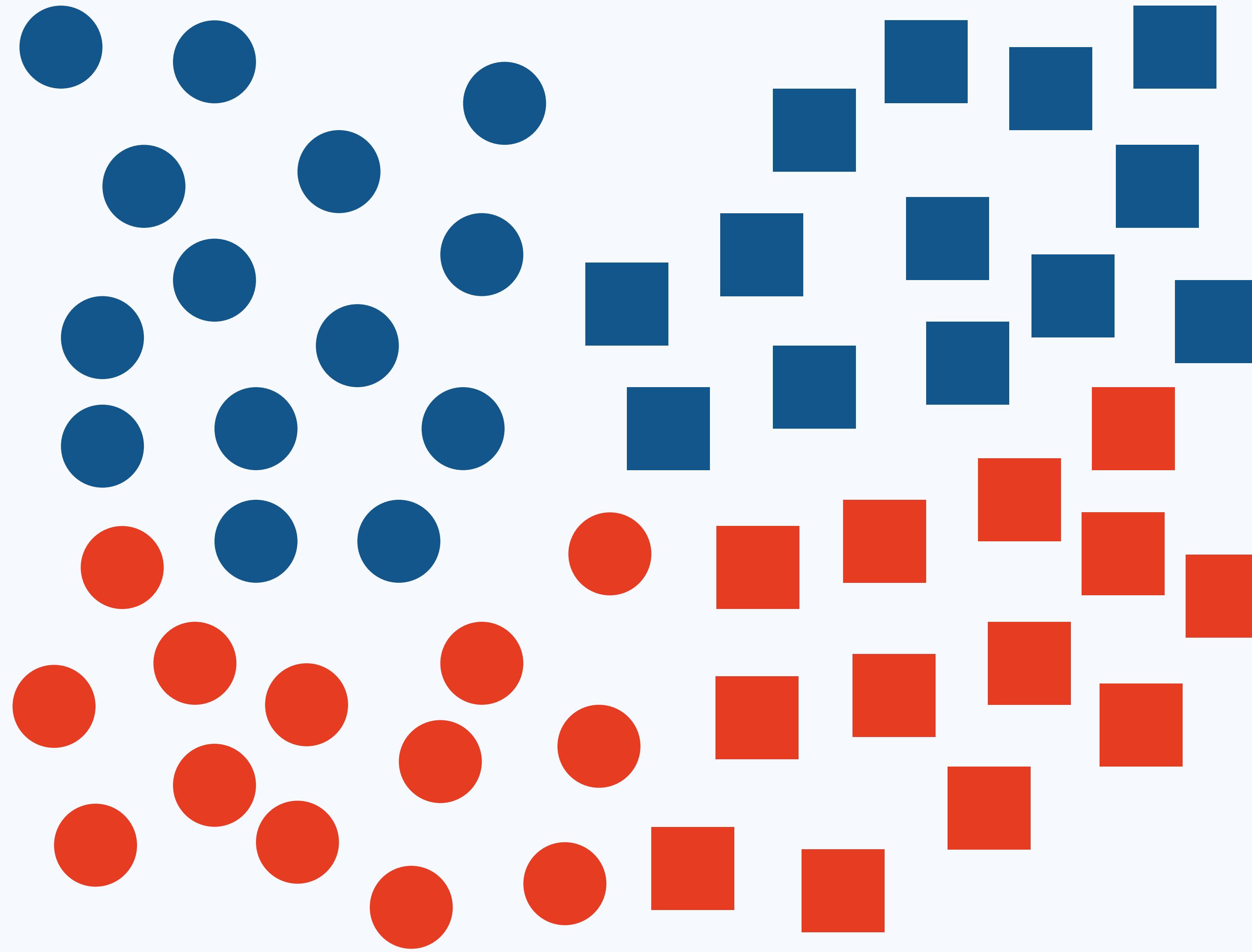
Benjamin Skov Kaas-Hansen (MD, MSc, PhD)

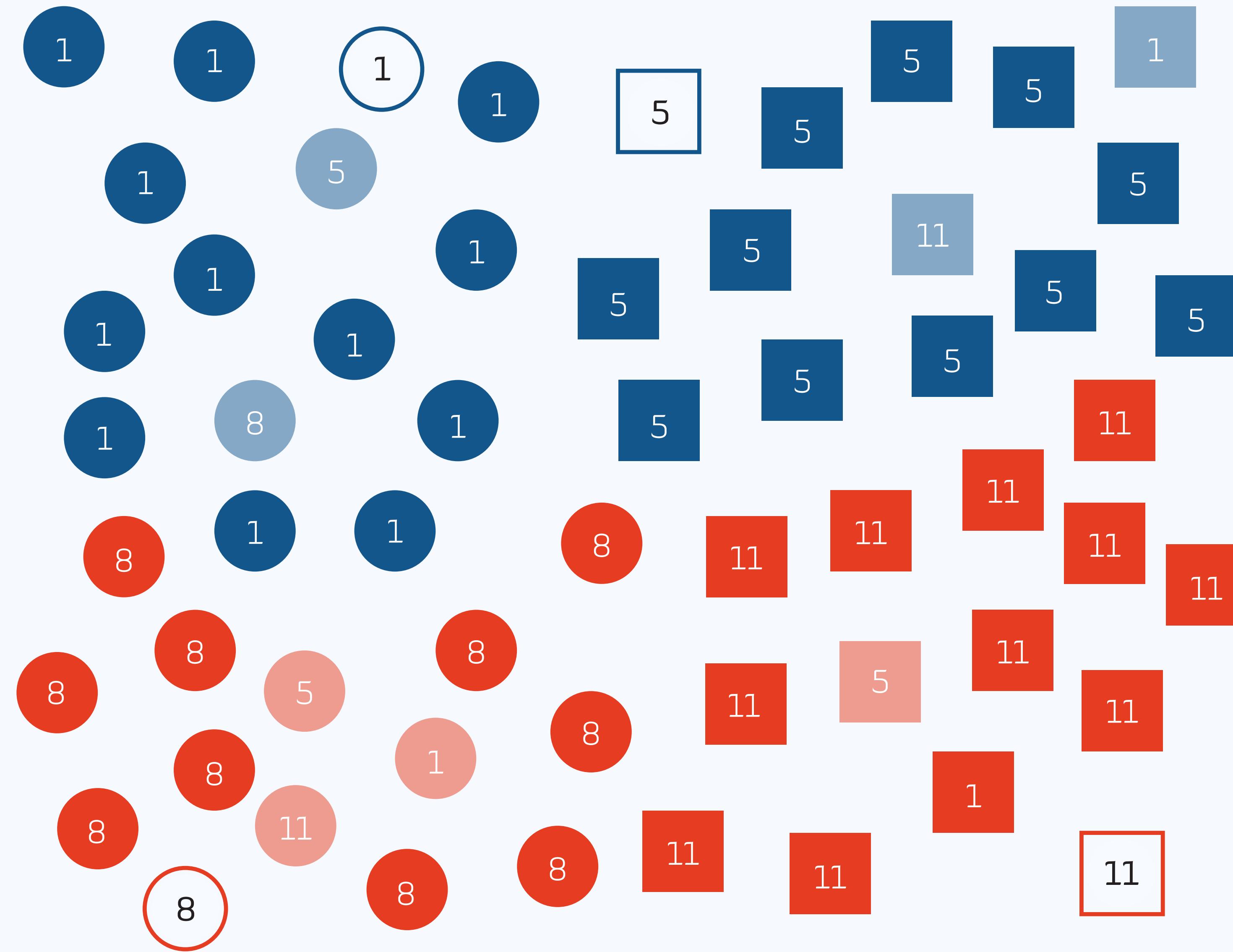
Postdoc på Afdeling for Intensiv Terapi, Rigshospitalet

DSKF Årsmøde – 8. oktober 2022

epiben@hey.com







GENKENDE HÅNDSKREVNE TAL



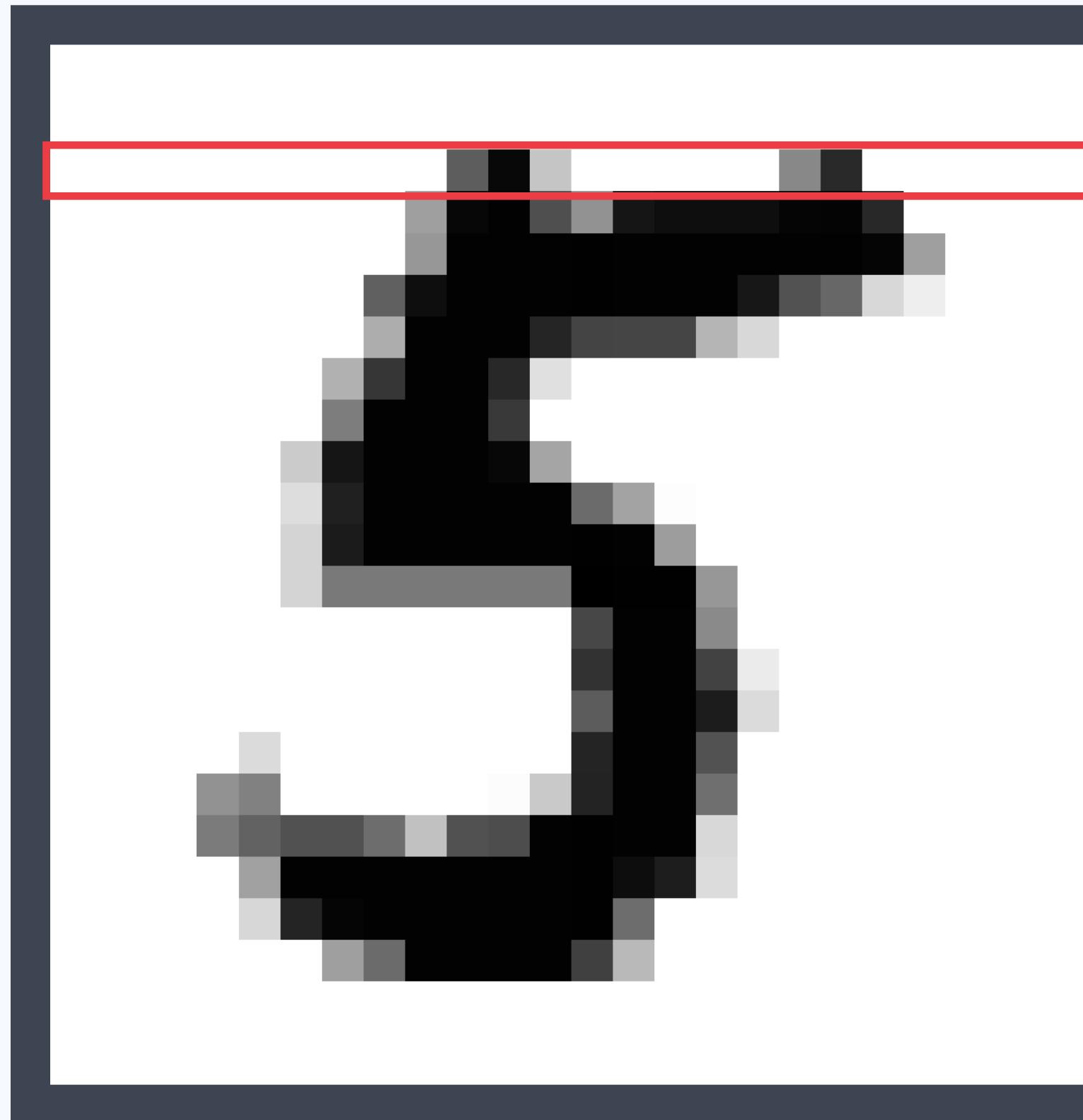
Hvert tal ~ 784 pixels (28×28)

Hver pixel ~ 1 feature

Feature-værdi ~ gråtone (0–1)

Alle rækker limes sammen til en vektor

GENKENDE HÅNDSKREVNE TAL



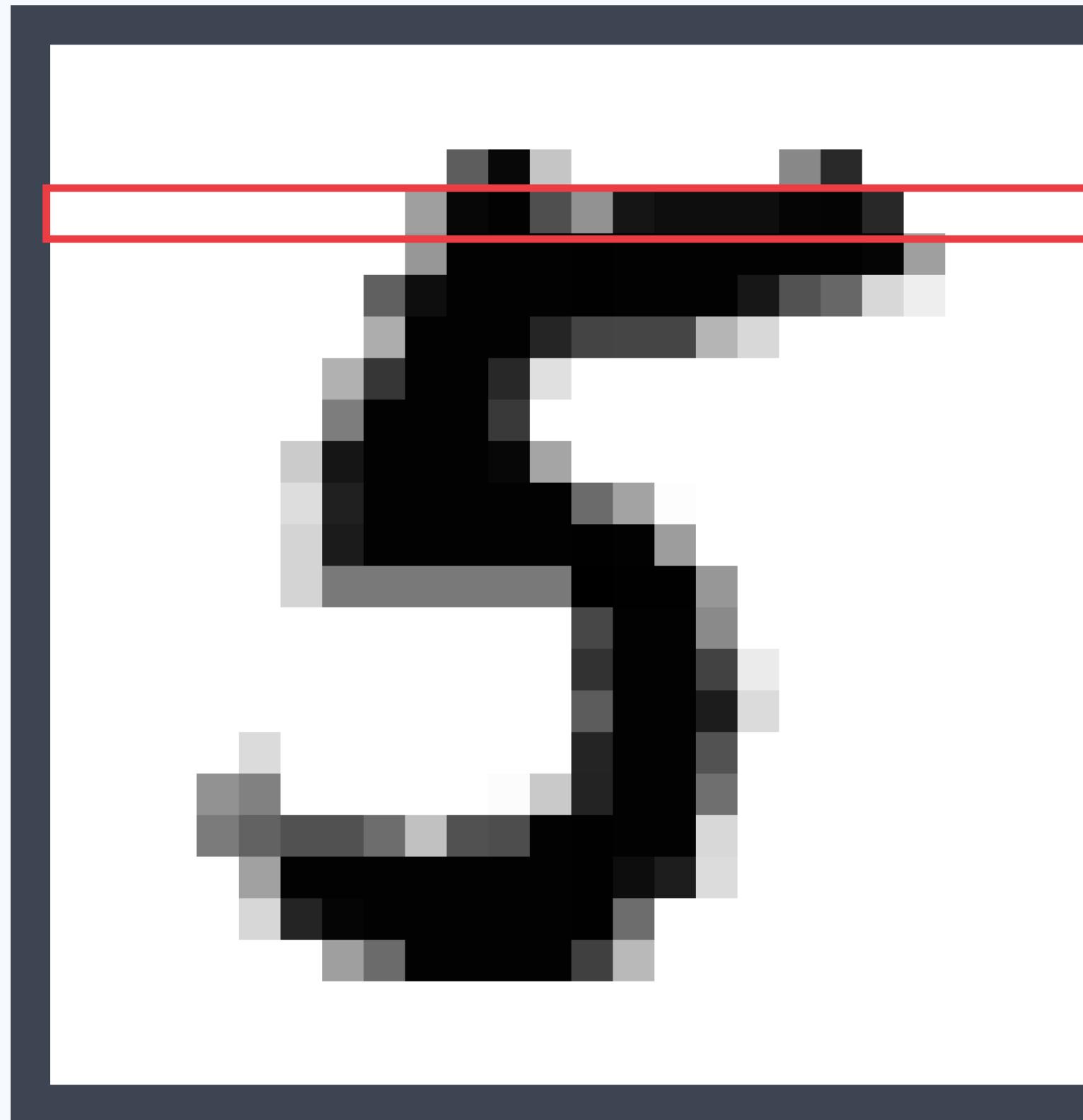
Hvert tal ~ 784 pixels (28×28)

Hver pixel ~ 1 feature

Feature-værdi ~ gråtone (0–1)

Alle rækker limes sammen til en vektor

GENKENDE HÅNDSKREVNE TAL

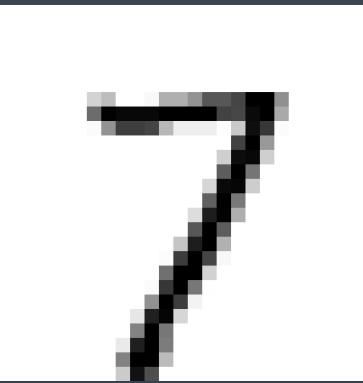


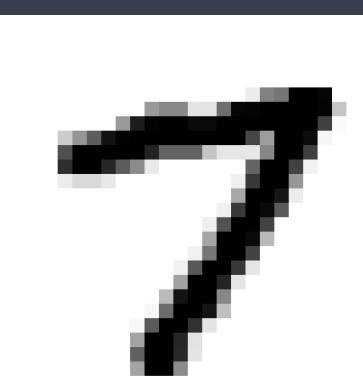
Hvert tal ~ 784 pixels (28×28)

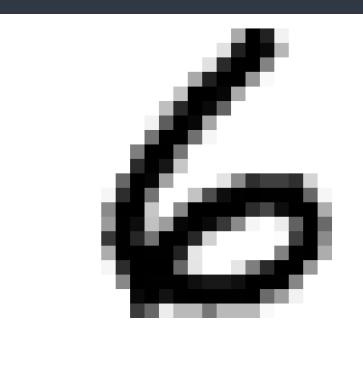
Hver pixel ~ 1 feature

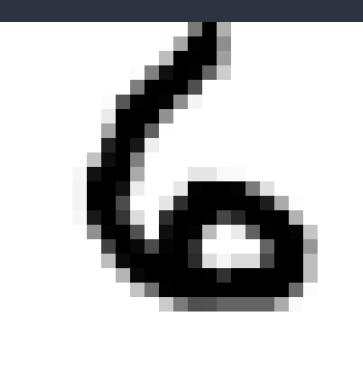
Feature-værdi ~ gråtone (0–1)

Alle rækker limes sammen til en vektor

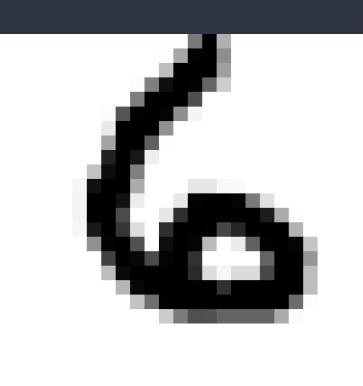
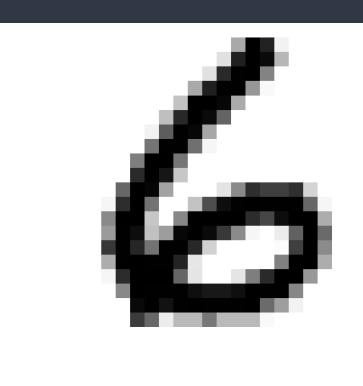
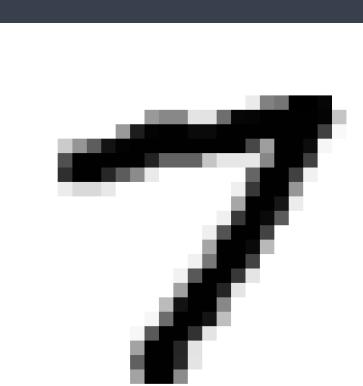
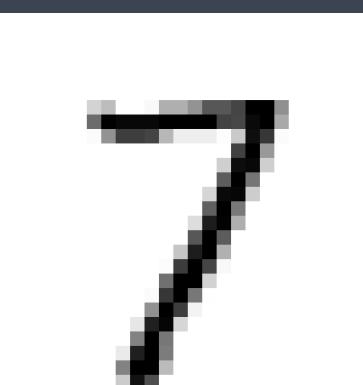






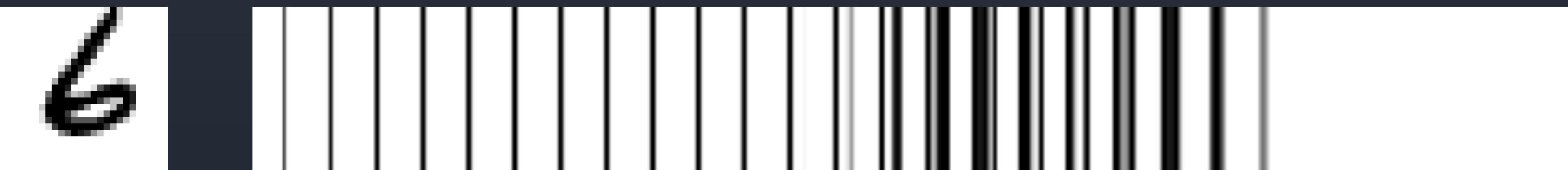


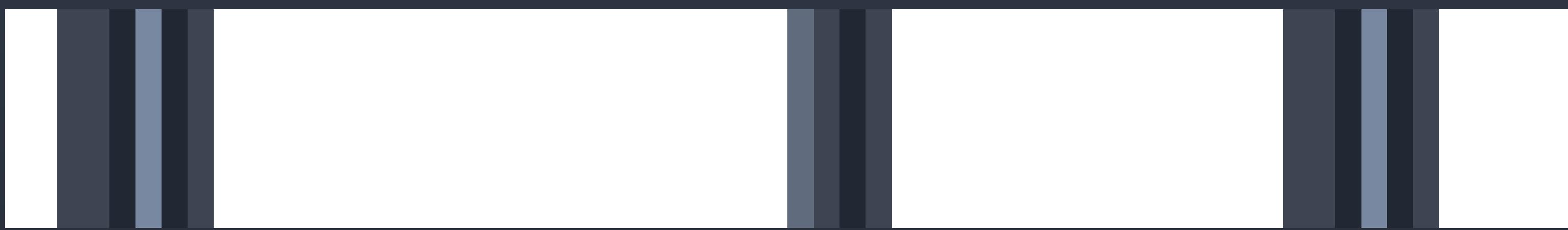
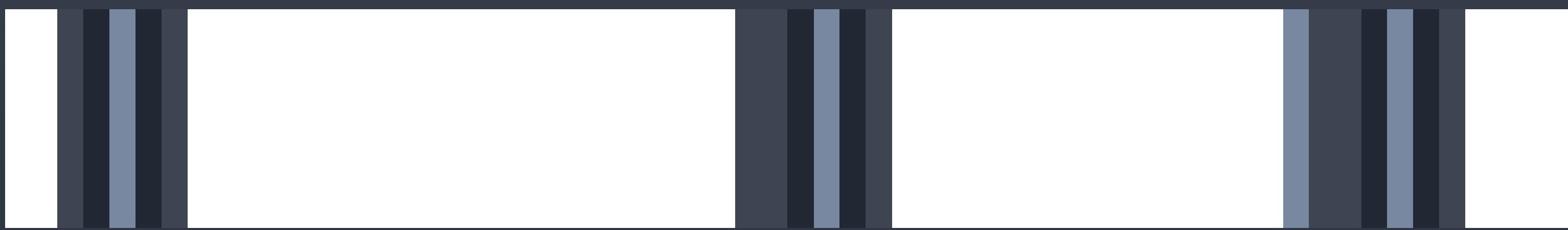
?



?

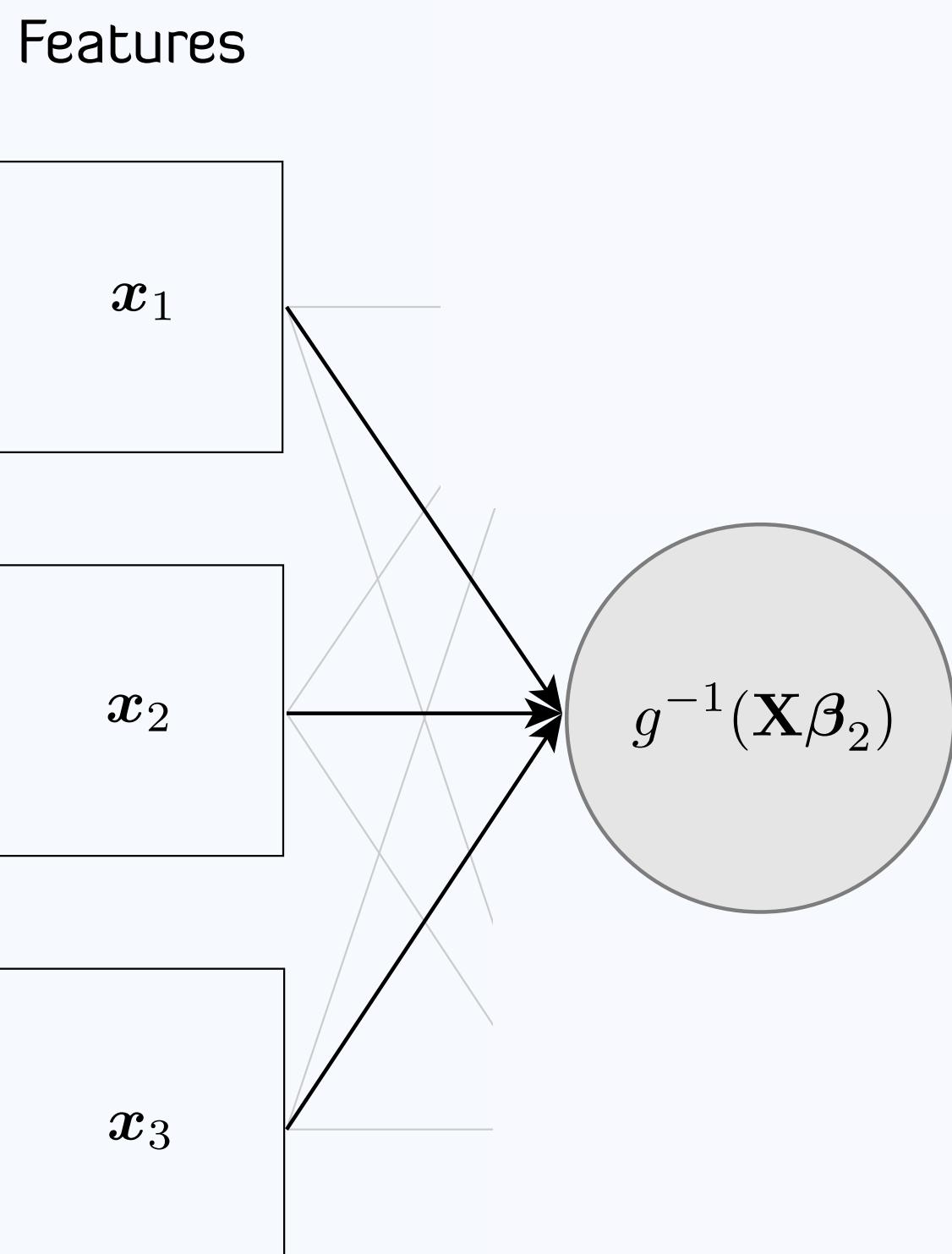




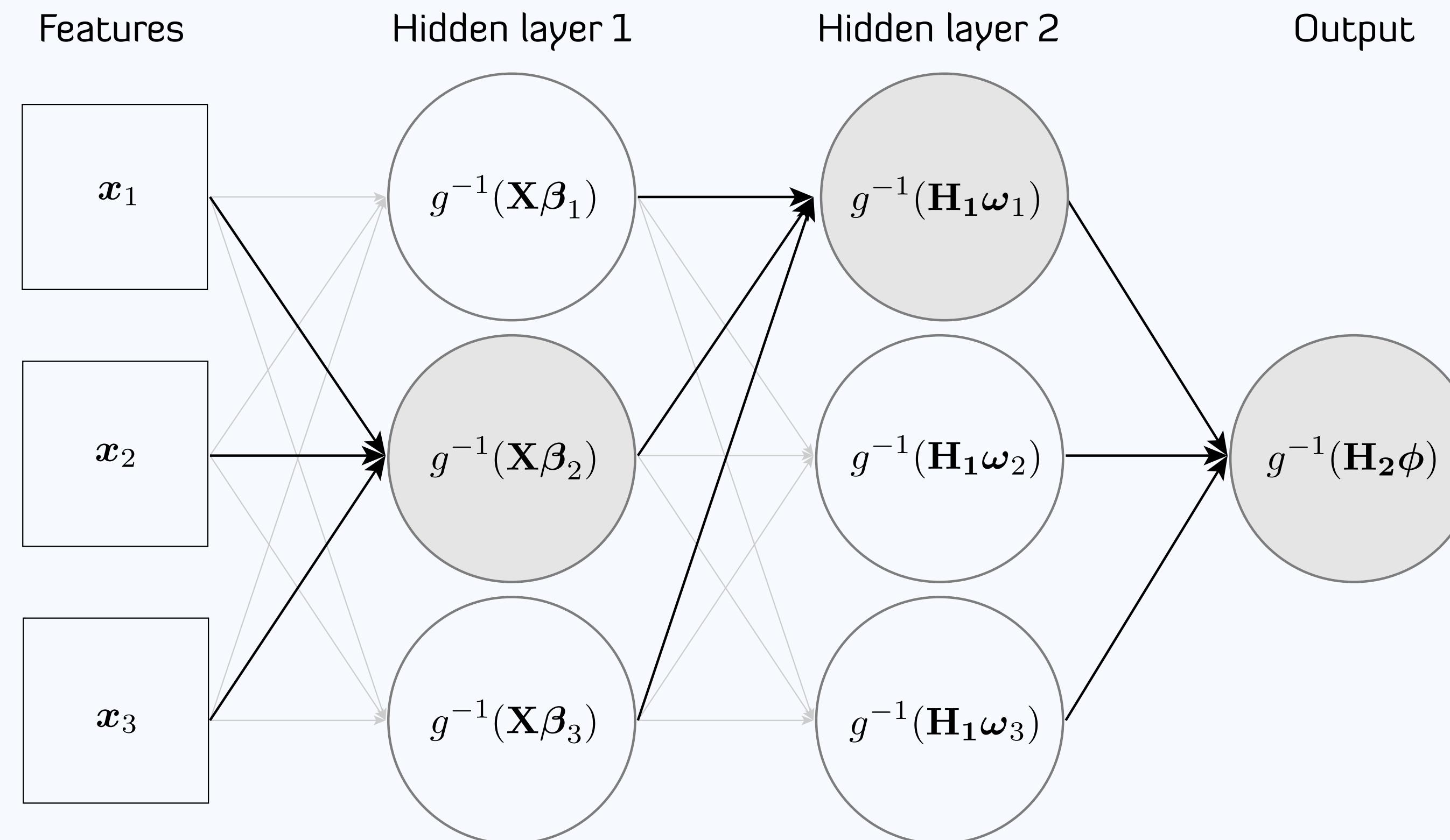




DEEP LEARNING 101



DEEP LEARNING 101



PREVALENCE AND ADVERSE OUTCOMES OF DRUG-DRUG INTERACTION

Causal inference about adverse outcomes of DDIs (part II)

Matching with high-dimensional propensity scores (lasso logistic regression)

843 features

Mortalitet og genindlæggelse: betinget Cox-regression

Length-of-stay: betinget Poisson-regression

RENAL DYSFUNCTION AND RISK OF INAPPROPRIATE DRUG DOSING

Prædiktionsstudy

>0 , ≥ 1 , ≥ 2 , ≥ 3 and ≥ 5 daily inappropriate doses af udvalgte stoffer

98 features (håndplukkede)

Ridge logistic regression vs. neural networks

xAI til at identificere *prediction drivers*

MACHINE LEARNING I JOURNALNOTATER

32-årig kvinde indbringes af Falck via egen læge pga. tiltagende åndeenød gennem 2-3 uger.

ANAMNESE: Opstartet p-pillebehandling for ca. 2 måneder siden. For 2-3 uger siden pludseligt opstået åndenød og intermitterende brystsmerter, ingen forværring eller bedring ved fysisk aktivitet eller hvile. Ingen hæmoptyse. Har ikke lagt mærke til rallende vejrtrækning. Patienten fortæller, at der har været flere tilfælde af blodpropper i familien, bl.a. døde patientens mor af en lungeemboli. Patienten egentligt velbefindende ved undersøgelsen.

OBJEKTIVT: VKO. Normale halsvener. St.c.: Regelmæssig aktion uden hørbare mislyde. St.p.: Vesikulær respiration bilateralt, ingen bilyde og normale lungegrænser. Abdomen fladt og uømt ved palpation, livlige tarmlyde. Overekstremitter og højre underekstremitet upåfaldende. Venstre læg rød og varm; lægømhed og popliteaømhed. Normale fodpulse bilateralt.

KONKLUSION: 32-årig kvinde, nyopstartet i p-pillebehandling nu med symptomer forenelige med DVT og lungeemboli. ABC-stabil ved undersøgelsen. Indlægges til videre behandling.

32-årig kvinde indbringes af Falck via egen læge pga. tiltagende **åndeenød** gennem 2-3 uger.

ANAMNESE: Opstartet p-pillebehandling for ca. 2 måneder siden. For 2-3 uger siden **pludseligt opstået åndenød** og **intermitterende brystsmerter**, ingen forværring eller bedring ved fysisk aktivitet eller hvile. Ingen hæmoptyse. Har ikke lagt mærke til rallende vejrtrækning. Patienten fortæller, at der har været flere tilfælde af blodpropper i familien, bl.a. døde patientens mor af en lungeemboli. Patienten egentligt velbefindende ved undersøgelsen.

OBJEKTIVT: VKO. Normale halsvener. St.c.: Regelmæssig aktion uden hørbare mislyde. St.p.: Vesikulær respiration bilateralt, ingen bilyde og normale lungegrænser. Abdomen fladt og uømt ved palpation, livlige tarmlyde. Overekstremitter og højre underekstremitet upåfaldende. Venstre læg **rød** og **varm**; **lægømhed** og **popliteaømhed**. Normale fodpulse bilateralt.

KONKLUSION: 32-årig kvinde, nyopstartet i p-pillebehandling nu med symptomer forenelige med **DVT** og **lungeemboli**. ABC-stabil ved undersøgelsen. Indlægges til videre behandling.

116 unikke ord

ONE-HOT-ENCODING AF TOKENS

Antal dimensioner =
antal unikke ord i corpus

Kan ikke håndtere stavefejl
og synonymer

	1	2	3	4	5	...	113	114	115	116	
åndeenød	1	0	0	0	0	...	0	0	0	0	1
åndenød	0	1	0	0	0	...	0	0	0	0	2
brystsmerter	0	0	1	0	0	...	0	0	0	0	3
rød	0	0	0	1	0	...	0	0	0	0	4
varm	0	0	0	0	1	...	0	0	0	0	5
...	:	:	:	:	:	⋮	0	0	0	0	⋮
lægømhed	0	0	0	0	0	0	1	0	0	0	113
popliteaømhed	0	0	0	0	0	0	0	1	0	0	114
DVT	0	0	0	0	0	0	0	0	1	0	115
lungeemboli	0	0	0	0	0	0	0	0	0	1	116

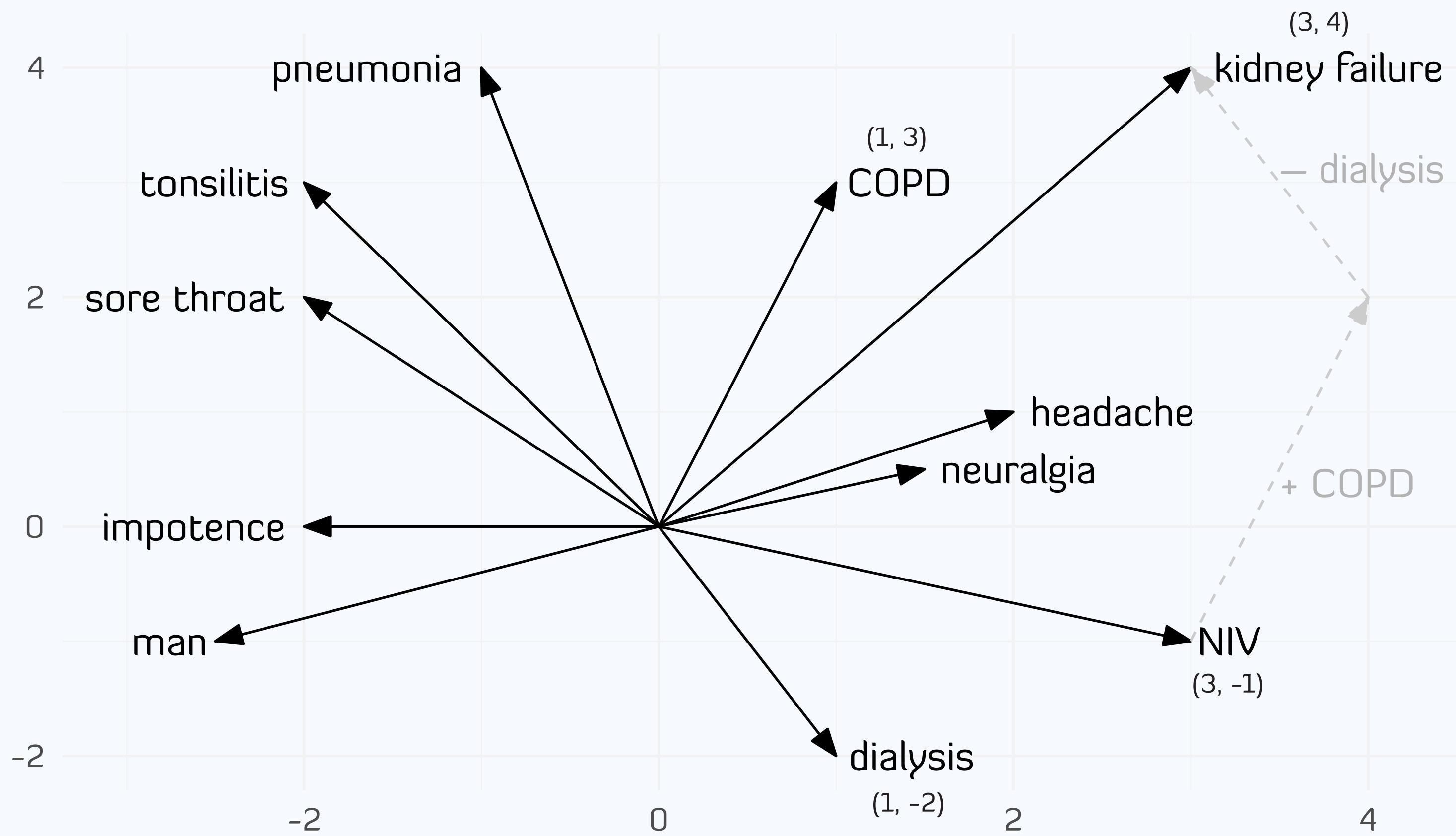
ONE-HOT- ENCODING AF TOKENS

Antal dimensioner =
antal unikke ord i corpus

Kan ikke håndtere stavefejl
og synonymer

	1	2	3	4	5	...	113	114	115	116	
1	0	0	0	0	0	...	0	0	0	0	1
0	1	0	0	0	0	...	0	0	0	0	2
0	0	1	0	0	0	...	0	0	0	0	3
0	0	0	1	0	0	...	0	0	0	0	4
0	0	0	0	1	0	...	0	0	0	0	5
:	:	:	:	:	:	:	0	0	0	0	:
0	0	0	0	0	0	0	1	0	0	0	113
0	0	0	0	0	0	0	0	1	0	0	114
0	0	0	0	0	0	0	0	0	1	0	115
0	0	0	0	0	0	0	0	0	0	1	116

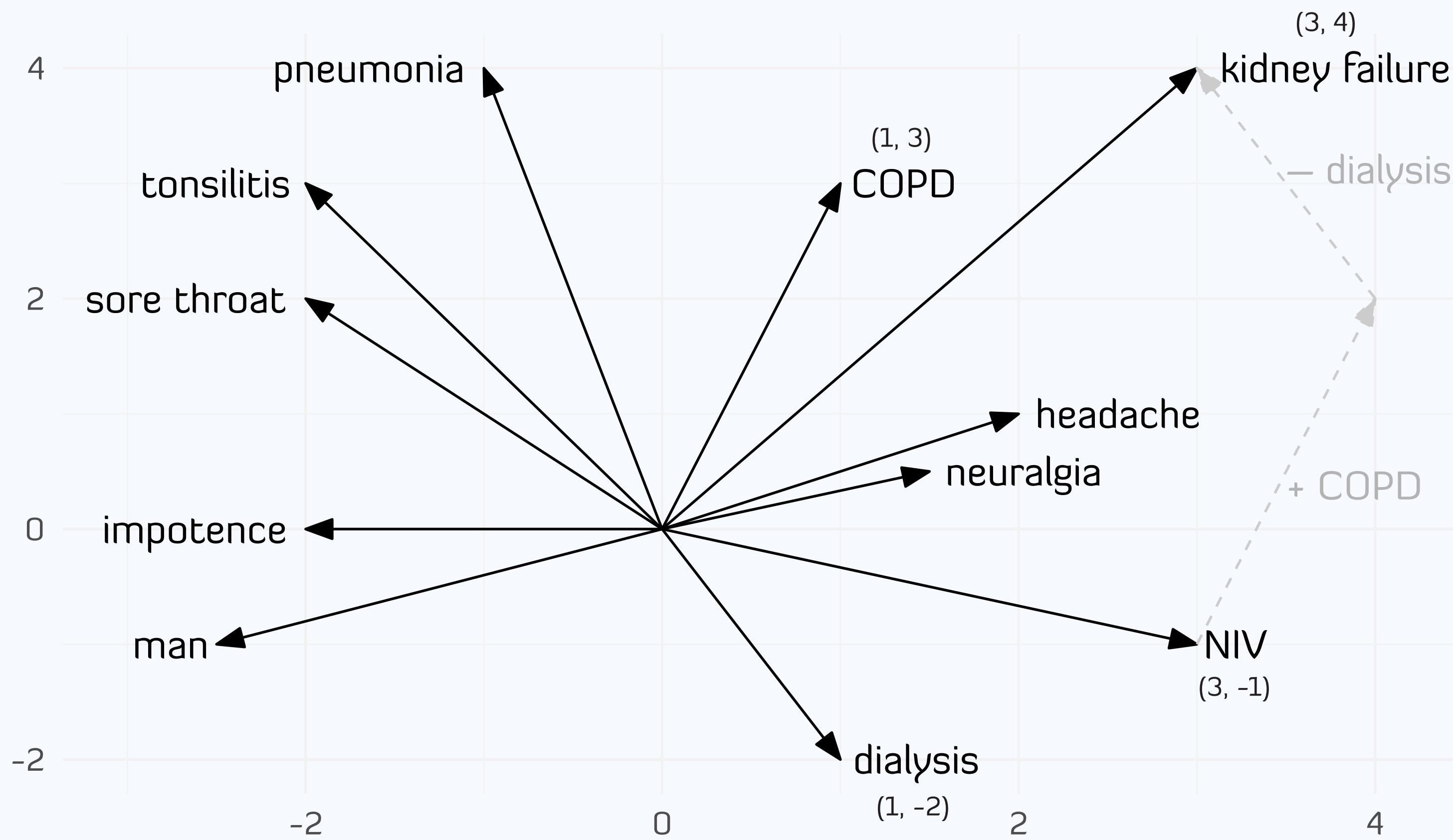
TOKEN EMBEDDINGS



NIV: non-invasive ventilation

COPD: chronic obstructive pulmonary disease

TOKEN EMBEDDINGS



$$NIV + COPD - dialysis$$

$$= \begin{array}{r} | \\ 3 \\ -1 \end{array} + \begin{array}{r} | \\ 1 \\ 3 \end{array} - \begin{array}{r} | \\ 1 \\ -2 \end{array}$$

$$= \begin{array}{r} | \\ 3 \\ 3 \end{array}$$

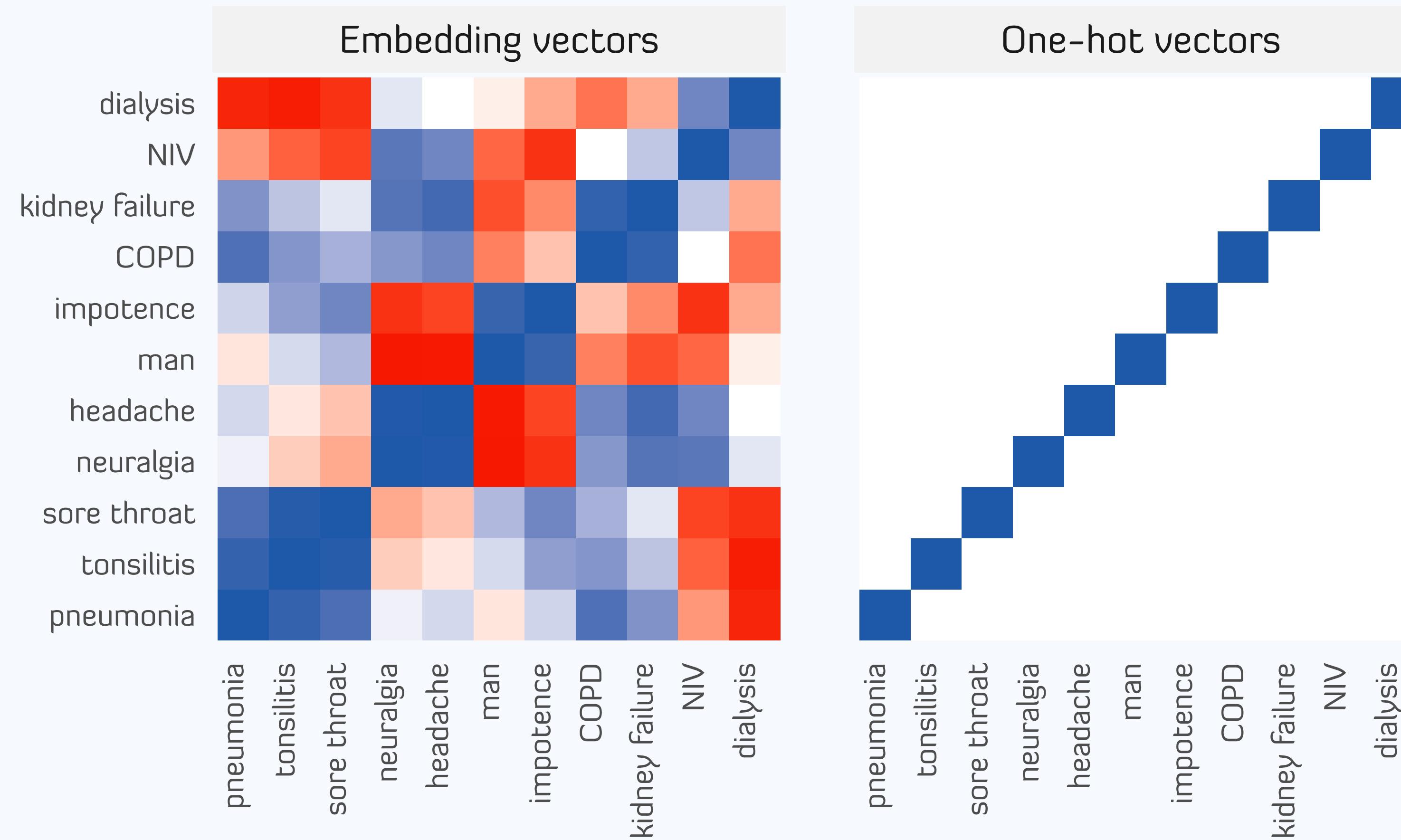
$$= \begin{array}{r} | \\ 4 \\ 4 \end{array}$$

$$= \text{kidney failure}$$

NIV: non-invasive ventilation

COPD: chronic obstructive pulmonary disease

VECTOR SIMILARITY



LANGUAGE-AGNOSTIC SAFETY SIGNAL DETECTION IN CLINICAL NOTES

Case-control-lignende design

Maj 2008 til juni 2016

Alle indlæggelser fra 500.000 tilfældigt udvalgte voksne

Doorstep medication profiles fra sygehusmedicindata

Alle notater fra første 48 timer af indlæggelse

fastText-embedding af tokens fra notater

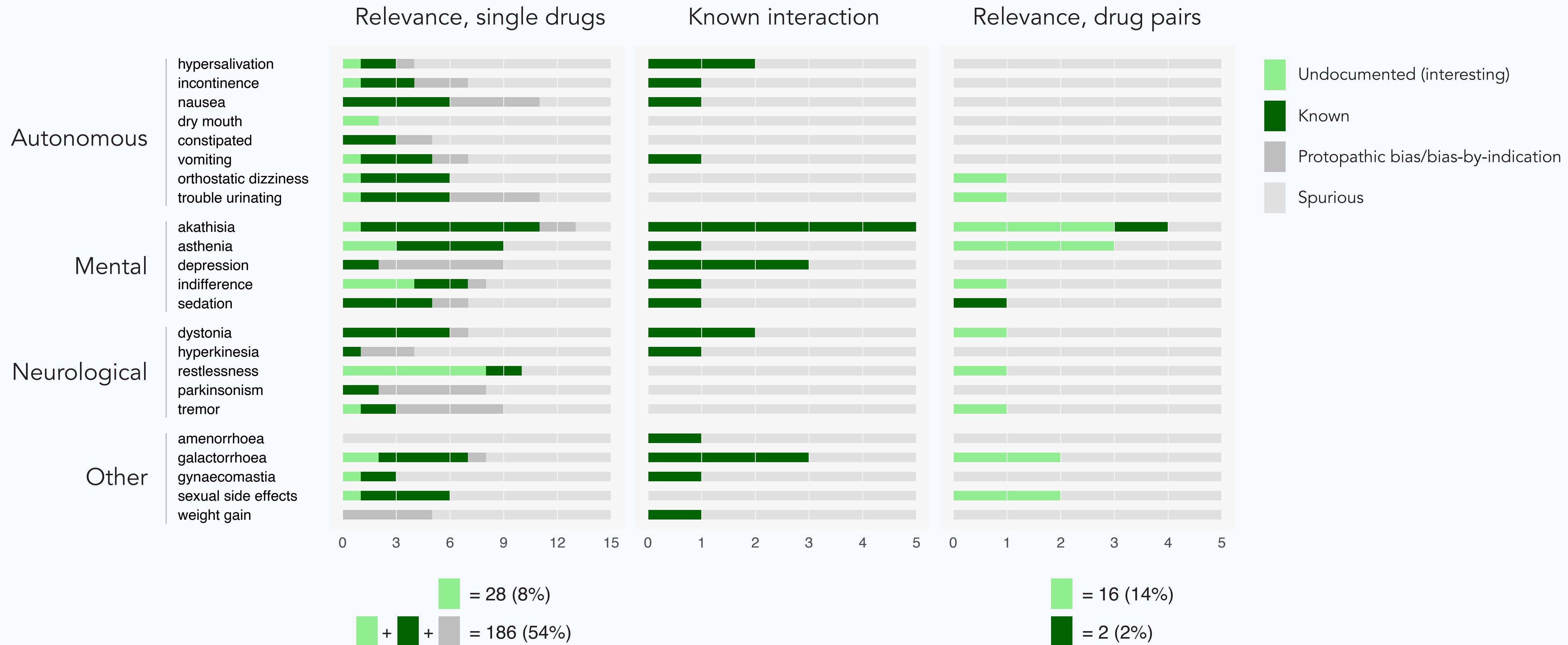
SUMMARY STATISTICS

Inpatient admissions	2,905,251 (54% of women)	←
Median age	58 years (IQR: 33-73)	
Clinical notes	10,788,259	
Doorstep drug prescriptions	13,740,564	
Admissions with ≥ 5 concurrent drugs	1,184,340 (41%)	
Tokens after pruning	179,441,739	→ per-admission median: 51 (IQR: 29-80)
Neural network models trained	10,270	←
Models with pertinent signals	3,945 (38%)	←

FINGERPRINTS (SINGLE DRUGS)



EVALUATION



KONKLUSION

End-to-end pipeline drug safety mining i fritekst og strukturerede medicindata

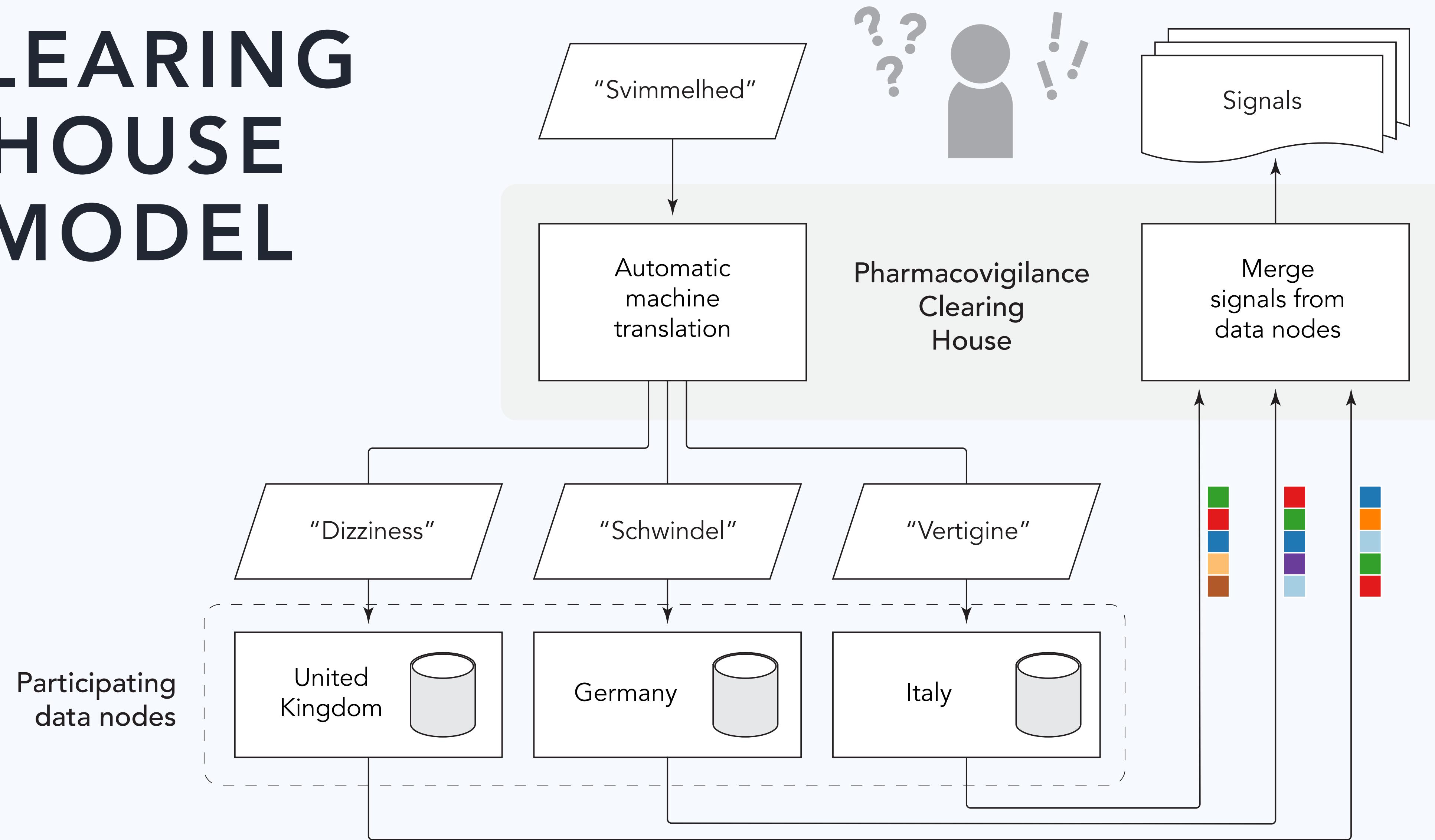
Okay hit-rate med fornuftige signaler

Kan afhjælpe problemer med metoder mest til engelsksproget tekst

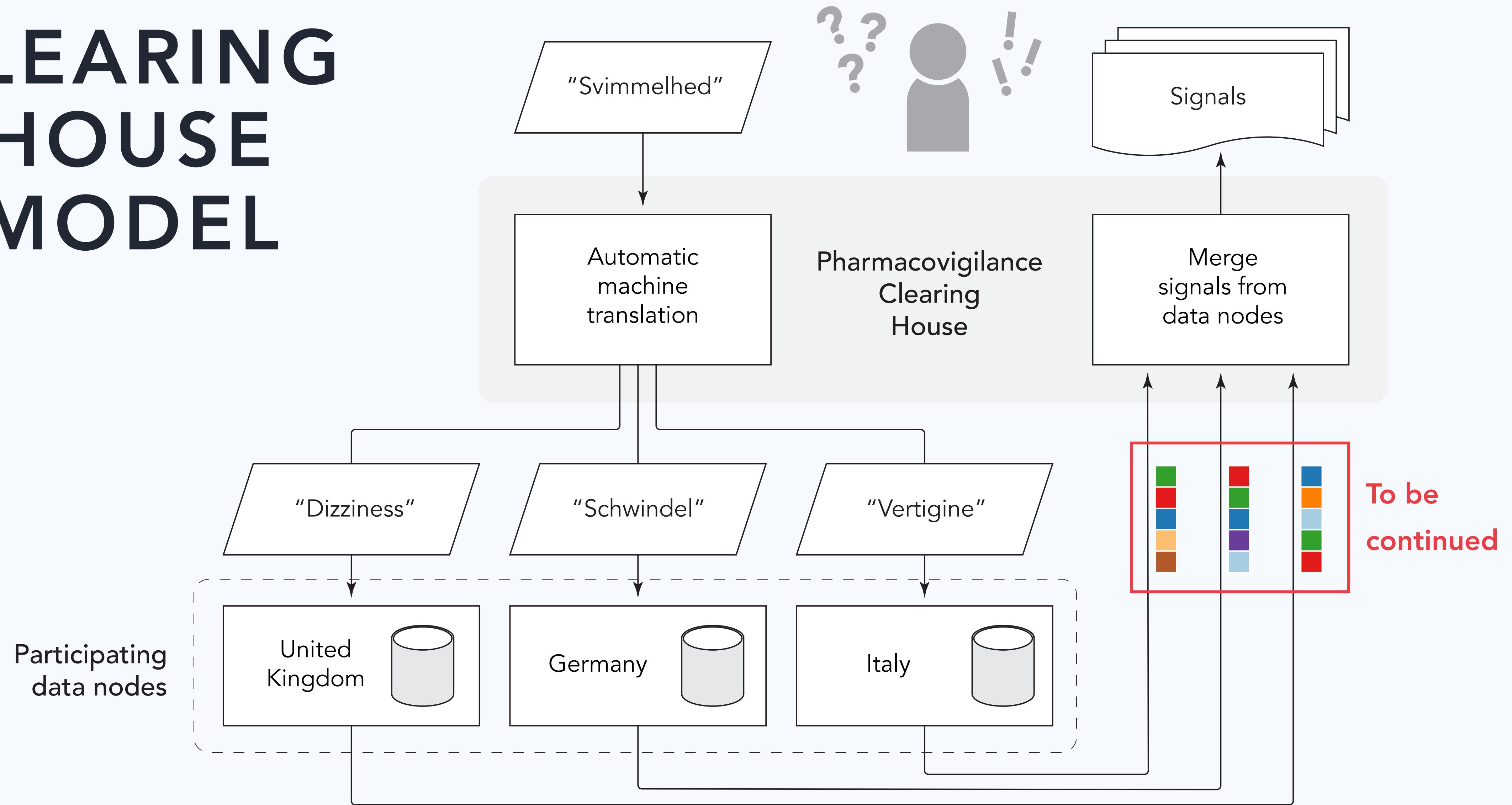
Embedding indeholder potentielt følsomme oplysninger => kan ikke deles

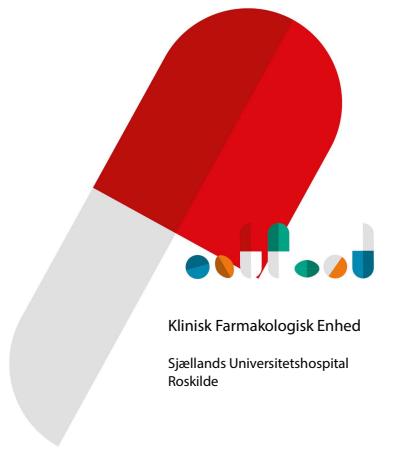
Behov for at kunne bruge data på tværs af sprog

CLEARING HOUSE MODEL

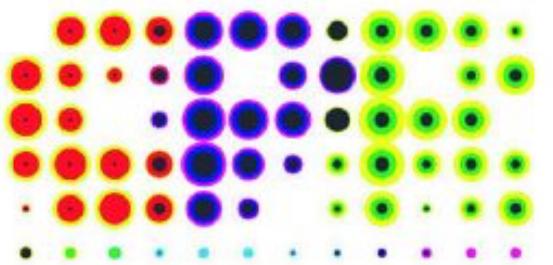


CLEARING HOUSE MODEL





Klinisk Farmakologisk Enhed
Sjællands Universitetshospital
Roskilde



The Novo Nordisk Foundation
Center for Protein Research



novo nordisk **fonden**

Innovation Fund Denmark

FINE

DATA TRANSFORMATION

Raw data
Normalisation
One-hot-encoding
Embedding
Mean encoding

PROCESS MODELLING

Generalised linear models
Penalised regression
Artificial neural networks
Tree-based models
Support vector machines

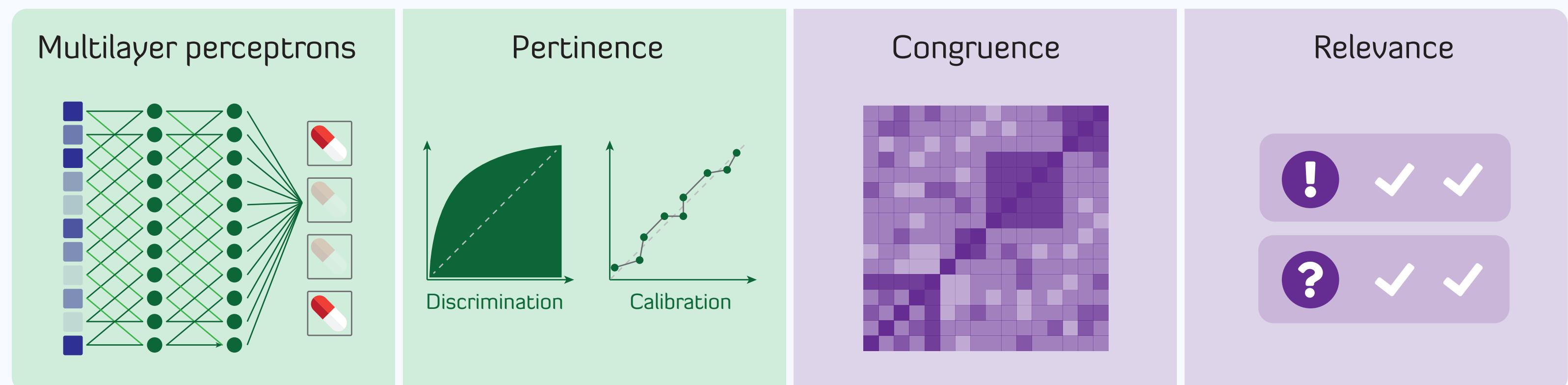
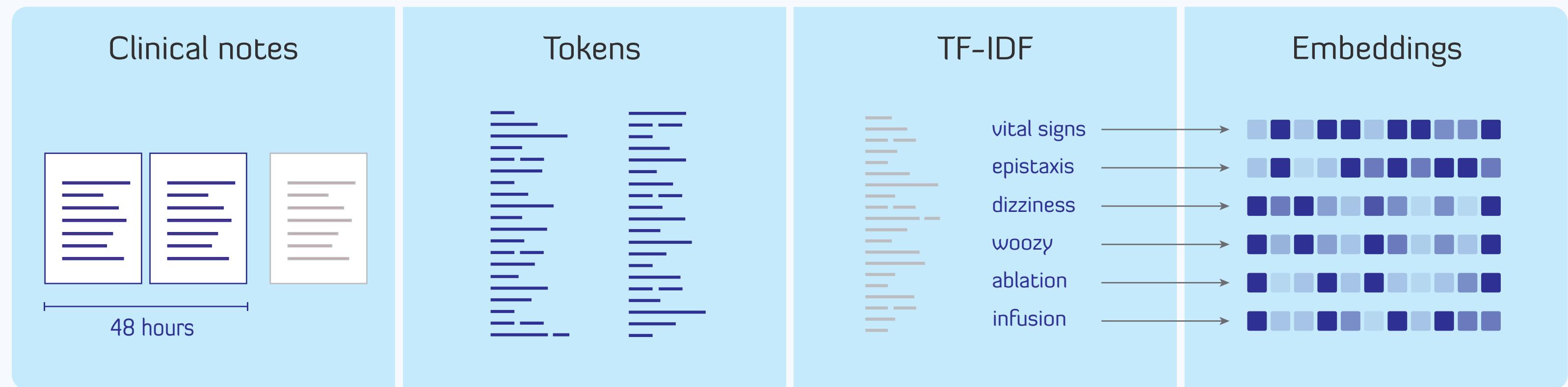
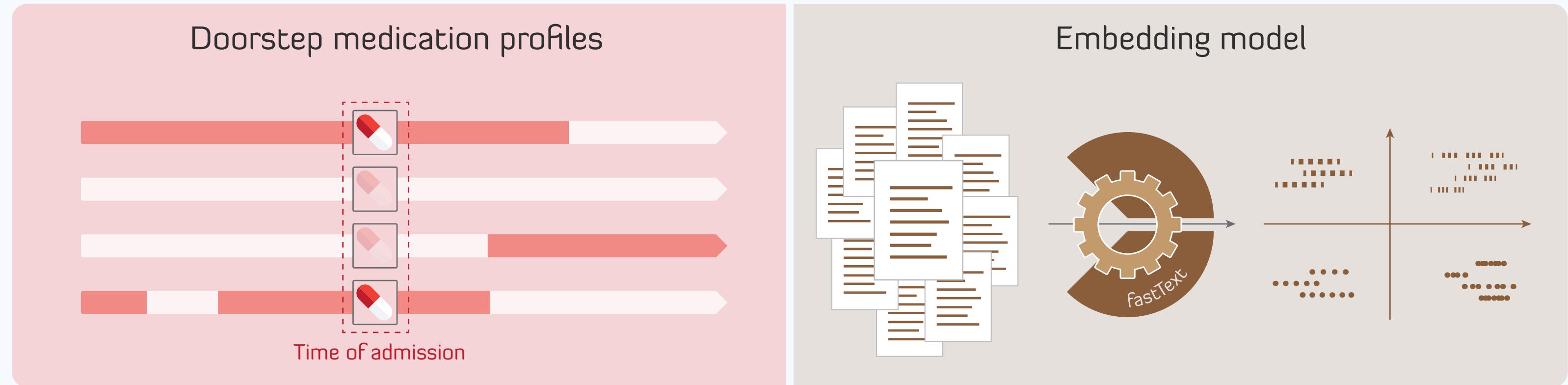
OUTCOME MODELLING

Binary (survival)
Continuous (biomarkers)
Discrete numeric (counts)
Multivariate (time-to-event)

DATA

Studieperiode	18. maj 2008 t.o.m. 30. juni 2016
Datakilder	EPM1+3 og OpusMedicin Journalnotater (kontinuationer)
Kriterier	Alle indlæggelser for 500.000 tilfældige voksne patienter
Referencesæt	ADRs: pro.medicin.dk DDIs: DrugBank, Danish Drug Interaction Database Manuelt review: UKU side effects rating scale

DESIGN



CONGRUENCE BETWEEN SIGNAL PROFILES

UKU domains (4) →
UKU terms (23) →
Synonyms (116) →



AUTOMATIC MACHINE TRANSLATION

