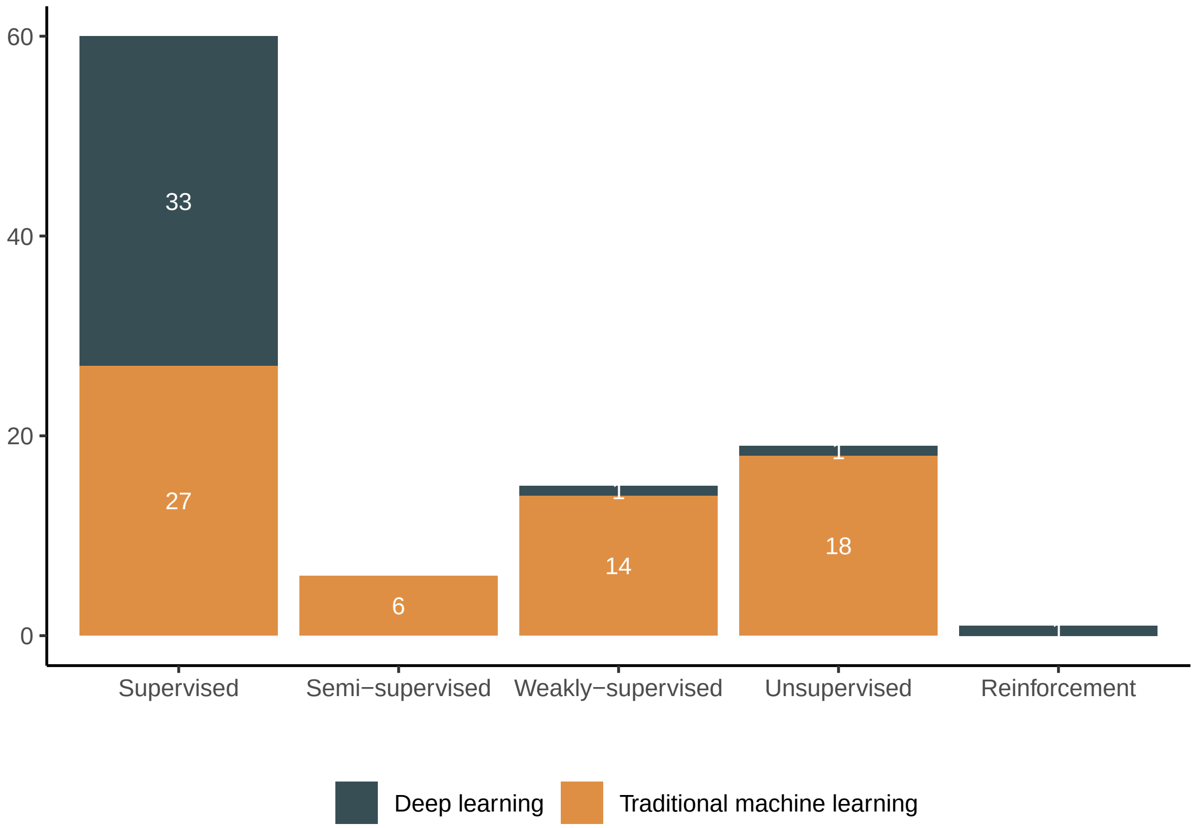
A summary of EHR-based phenotyping article annotation

Siyue Yang, Jessica Gronsbell

05/25/2022

Contents

1 Overview



1.1 Traditional ML method

Table 1: Common traditional machine learning methods (Count > 1)

ML	Traditional ML method	Count
Supervised	Random forest	14
Supervised	Logistic regression	11
Supervised	SVM	11
Supervised	L1 logistic regression	8
Unsupervised	LDA	5
Supervised	Decision trees	4
Supervised	XGBoost	4
Unsupervised	K-means	4
Supervised	Naive Bayes	3
Weakly-supervised	PheNorm	3
Weakly-supervised	MAP	2
Weakly-supervised	Random forest	2
Unsupervised	UPGMA Hierarchical clustering	2

```
## [1] "There are 18 papers using multiple traditional machine learning methods"
```

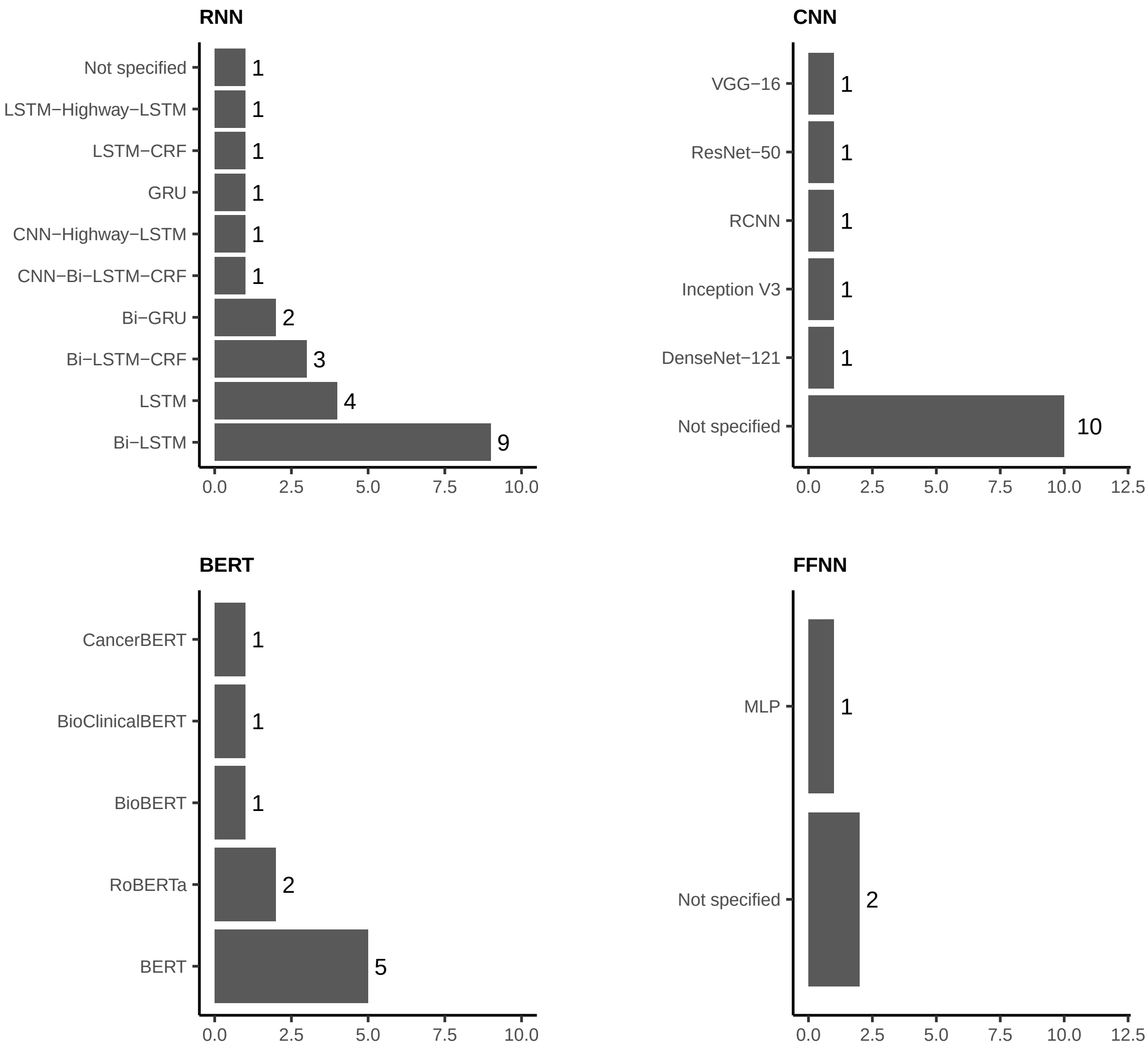
1.2 DL method

Table 2: Deep supervised learning methods

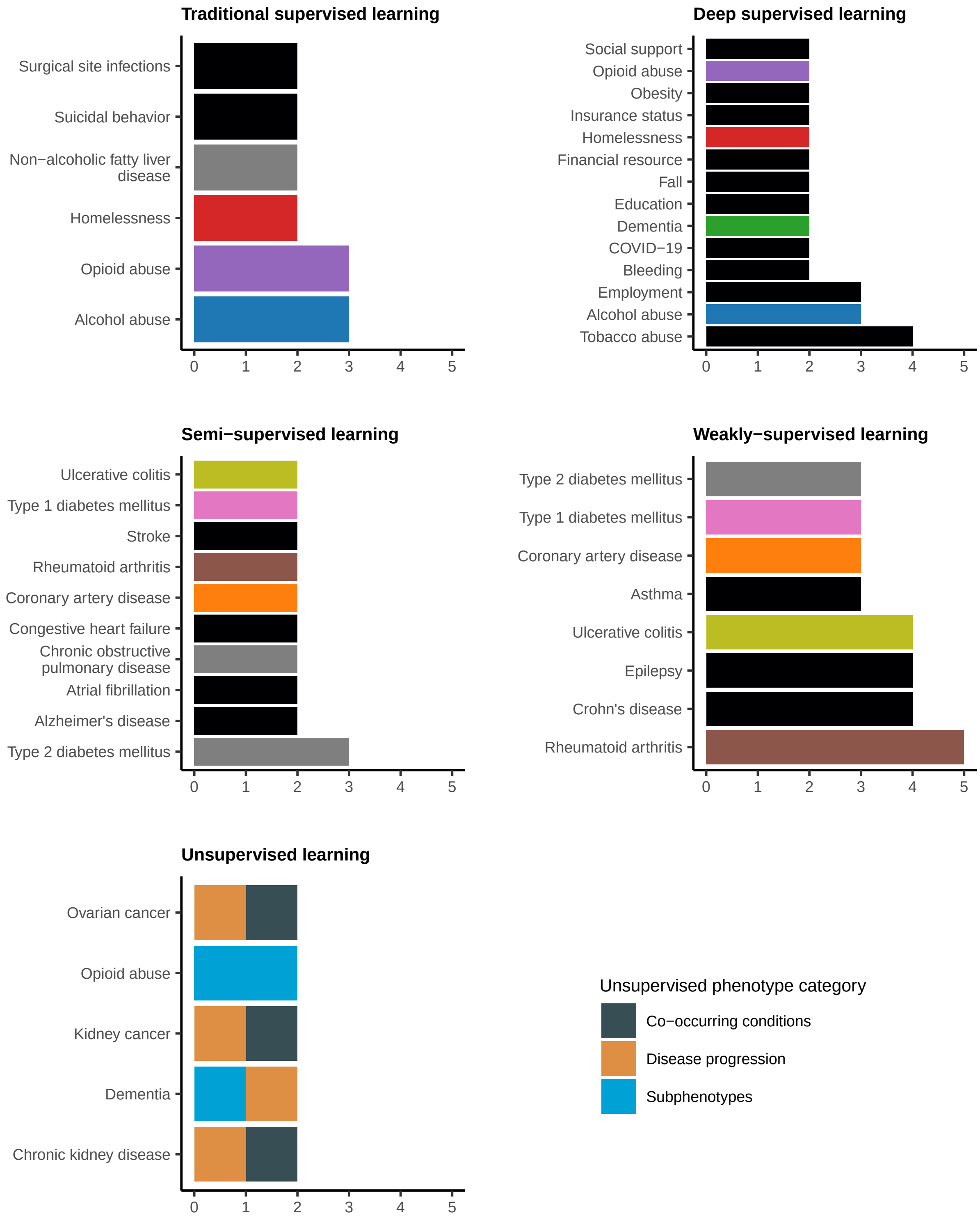
DL method	ML	Count
RNN	Supervised	18
CNN	Supervised	12
BERT	Supervised	7
FFNN	Supervised	3

```
## [1] "There are 5 papers using multiple deep learning methods"
```

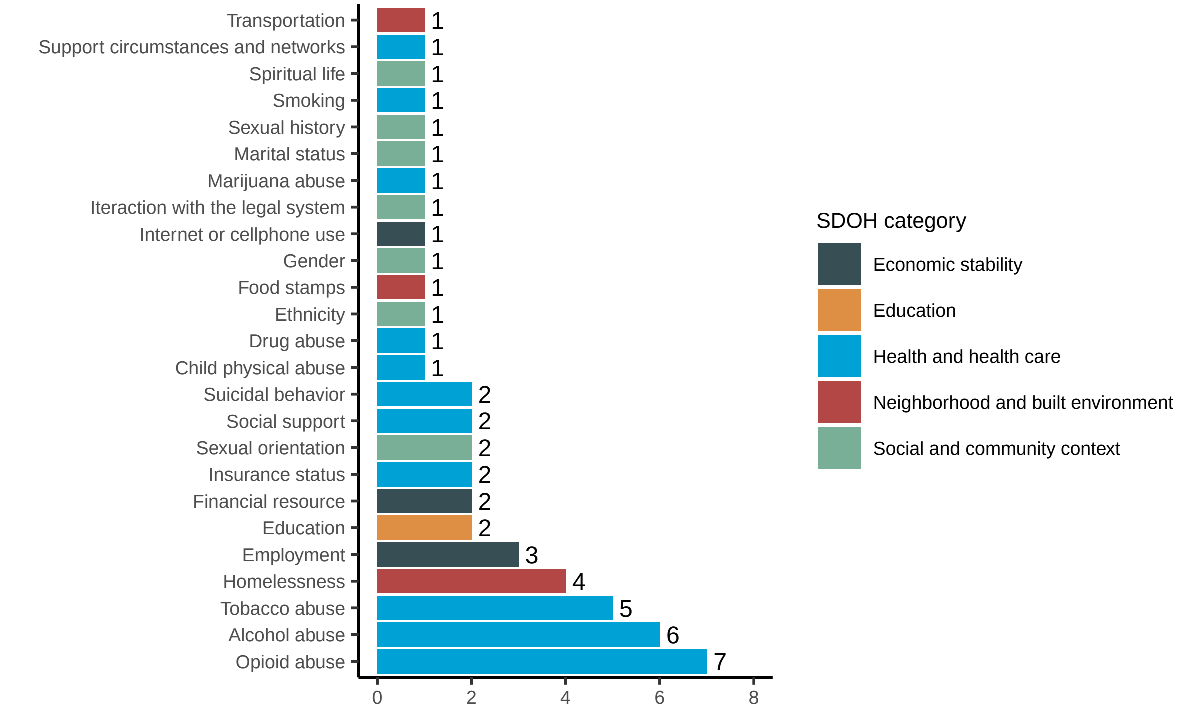
1.2.1 Deep neural network variants

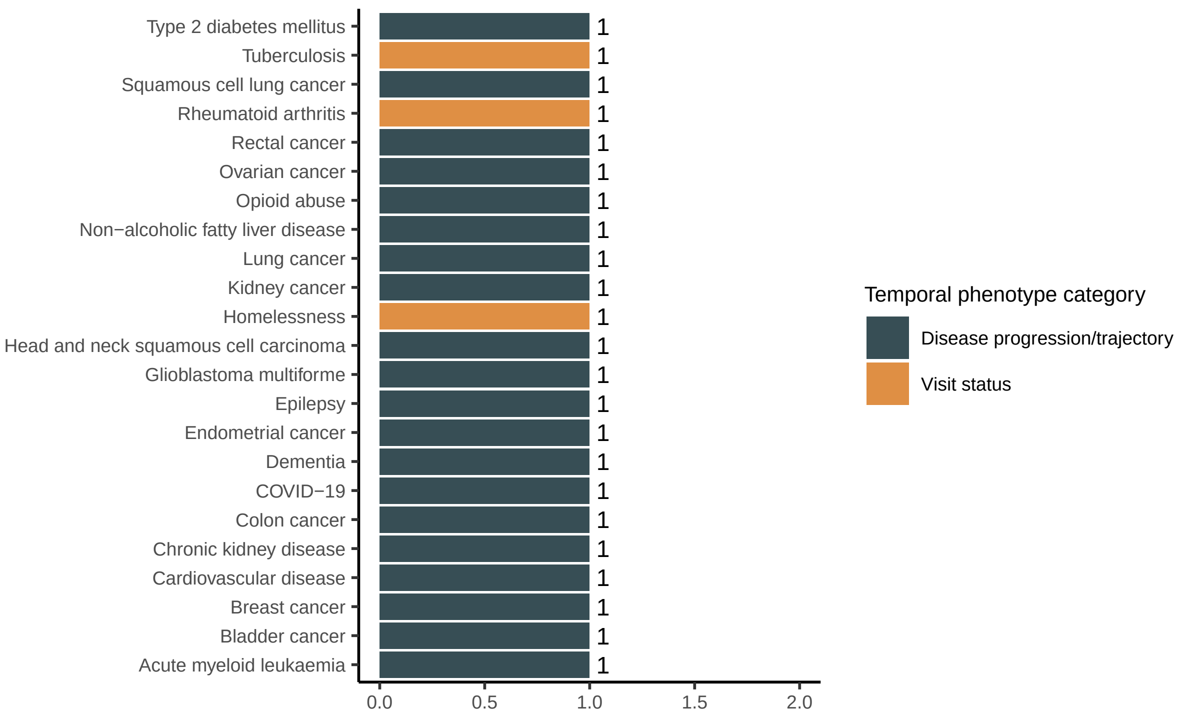


2 Phenotype


Traditional supervised learning
Surgical site infections
Suicidal behavior
Non−alcoholic fatty liver disease
Homelessness
Opioid abuse
Alcohol abuse
0 1 2 3 4 5
Deep supervised learning
Social support
Opioid abuse
Obesity
Insurance status
Homelessness
Financial resource
Fall
Education
Dementia
COVID−19
Bleeding
Employment
Alcohol abuse
Tobacco abuse
0 1 2 3 4 5
Semi−supervised learning
Ulcerative colitis
Type 1 diabetes mellitus
Stroke
Rheumatoid arthritis
Coronary artery disease
Congestive heart failure
Chronic obstructive pulmonary disease
Atrial fibrillation
Alzheimer's disease
Type 2 diabetes mellitus
0 1 2 3 4 5
Weakly−supervised learning
Type 2 diabetes mellitus
Type 1 diabetes mellitus
Coronary artery disease
Asthma
Ulcerative colitis
Epilepsy
Crohn's disease
Rheumatoid arthritis
0 1 2 3 4 5
Unsupervised learning
Ovarian cancer
Opioid abuse
Kidney cancer
Dementia
Chronic kidney disease
0 1 2 3 4 5
Unsupervised phenotype category
Co−occurring conditions
Disease progression
Subphenotypes

2.1 More nuanced phenotype

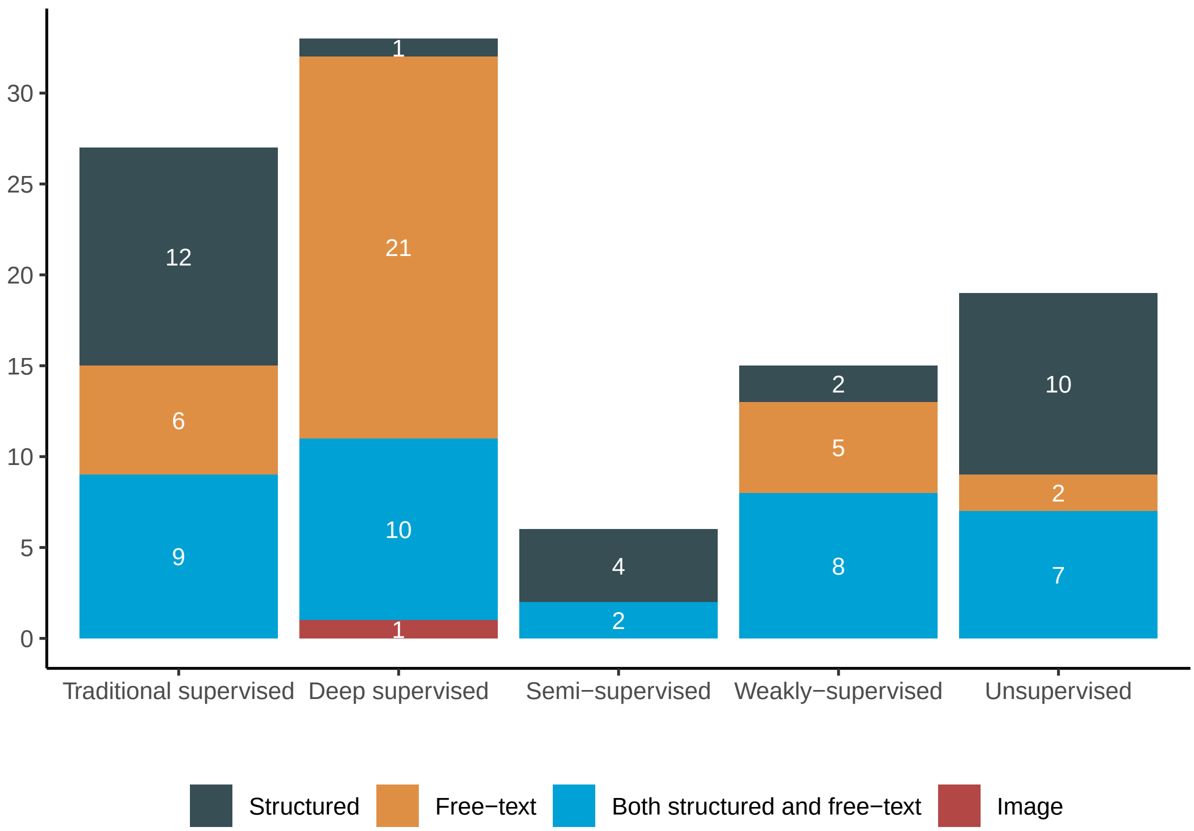




	Total number of papers	Used free-text	Used NLP software	Used competi-tion data	Used multisite data	Used open data	Used private single-site data	Compared to rule-based algo-rithms	Comapred to tradi-tional ML	Reported patient demo-graphic	Released open code
TSL	27	15	14	3	2	4	22	10	0	13	4
DSL	33	31	18	11	9	19	13	2	19	9	8
SSL	6	2	1	0	0	0	6	1	0	3	0
WSL	15	13	10	0	4	2	10	8	1	4	3
USL	19	9	4	0	3	3	13	0	0	15	4
Total	101	70	47	14	18	29	64	21	21	45	20

3 Data source

3.1 Summary

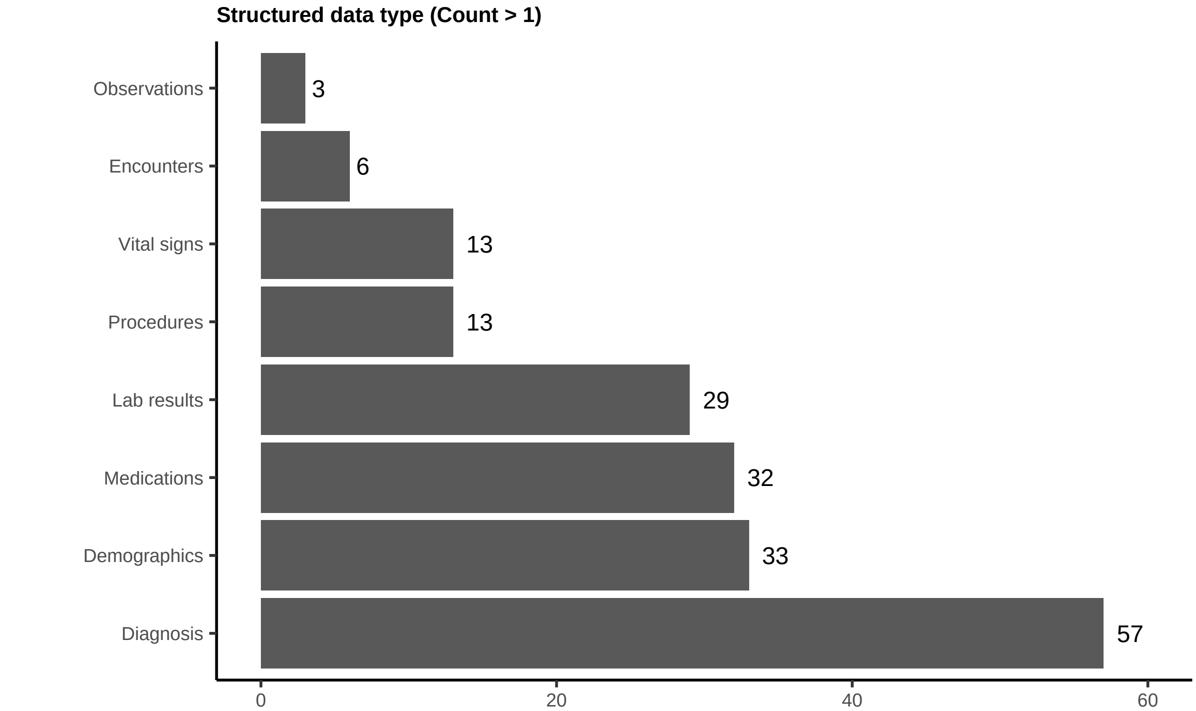


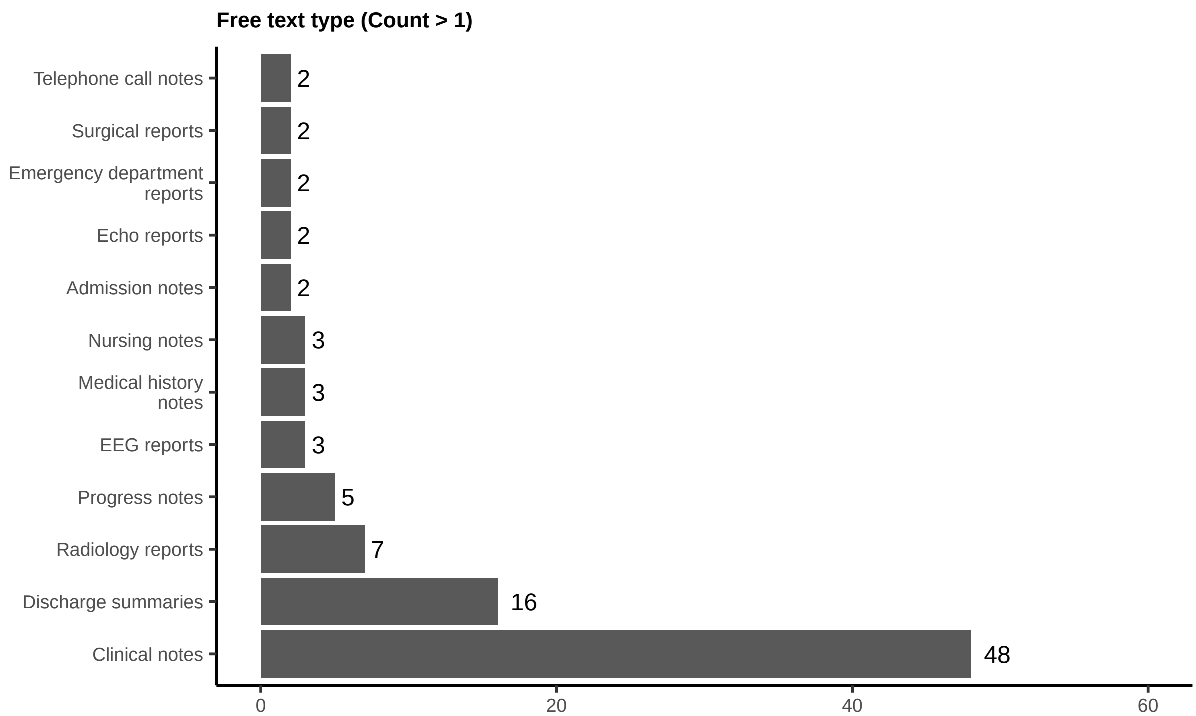
TSL = Traditional supervised learning. DSL = Deep supervised learning. DRL = Reinforcement deep learning. SSL = Semi-supervised learning. WSL = Weakly-supervised learning. US = Unsupervised learning.

3.2 Structured and unstructured data type

```
## [1] "There are 50 papers using multiple structured data type"
```

```
## [1] "There are 15 papers using multiple unstructured data type"
```


Structured data type (Count > 1)
Observations
Encounters
Vital signs
Procedures
Lab results
Medications
Demographics
Diagnosis
3
6
13
13
29
32
33
57
0
20
40
60


Free text type (Count > 1)
Telephone call notes
Surgical reports
Emergency department reports
Echo reports
Admission notes
Nursing notes
Medical history notes
EEG reports
Progress notes
Radiology reports
Discharge summaries
Clinical notes
2
2
2
2
2
3
3
3
5
7
16
48
0
20
40
60

Data source	Supervised Deep learning	Supervised Traditional machine learning	Weakly-supervised Deep learning	Weakly-supervised Traditional machine learning	Unsupervised Traditional machine learning	Reinforcement Deep learning	Count
MIMIC-III database	14	1	1	1	3	1	21

3.3 Openly-available data

```
## [1] "There are 2 papers using multiple Competition data"
```

Competition data name	Supervised Traditional machine learning	Supervised Deep learning	Count
2018 n2c2 track 2	0	6	6
2018 n2c2 track 1	1	3	4
TRECMED 2011	1	1	2
TRECMED 2012	1	1	2

Data source	Count
MIMIC-III database	21
MTSamples database	1

```
## [1] "There are 1 papers using multiple Openly data"
```

NLP software	Supervised Deep learning	Weakly-supervised Traditional machine learning	Supervised Traditional machine learning	Semi-supervised Traditional machine learning	Unsupervised Traditional machine learning	Count
cTAKES	8	0	8	1	2	19
NegEx	0	2	3	0	1	6
NILE	0	5	1	0	0	6
NLTK	4	0	0	0	1	5
MetaMap	1	0	3	0	0	4
Stanford CoreNLP	2	0	0	0	0	2

4 NLP software

```
## [1] "There are 7 papers using multiple NLP software"
```

5 Emebddings

Embeddings were only used in deep supervised articles.

Embedding training data	Count
Unstructured EHR	11
Biomedical literature	10
MIMIC-III database (internal)	7
MIMIC-III database (external)	6
Wikipedia	6
Structured EHR	2

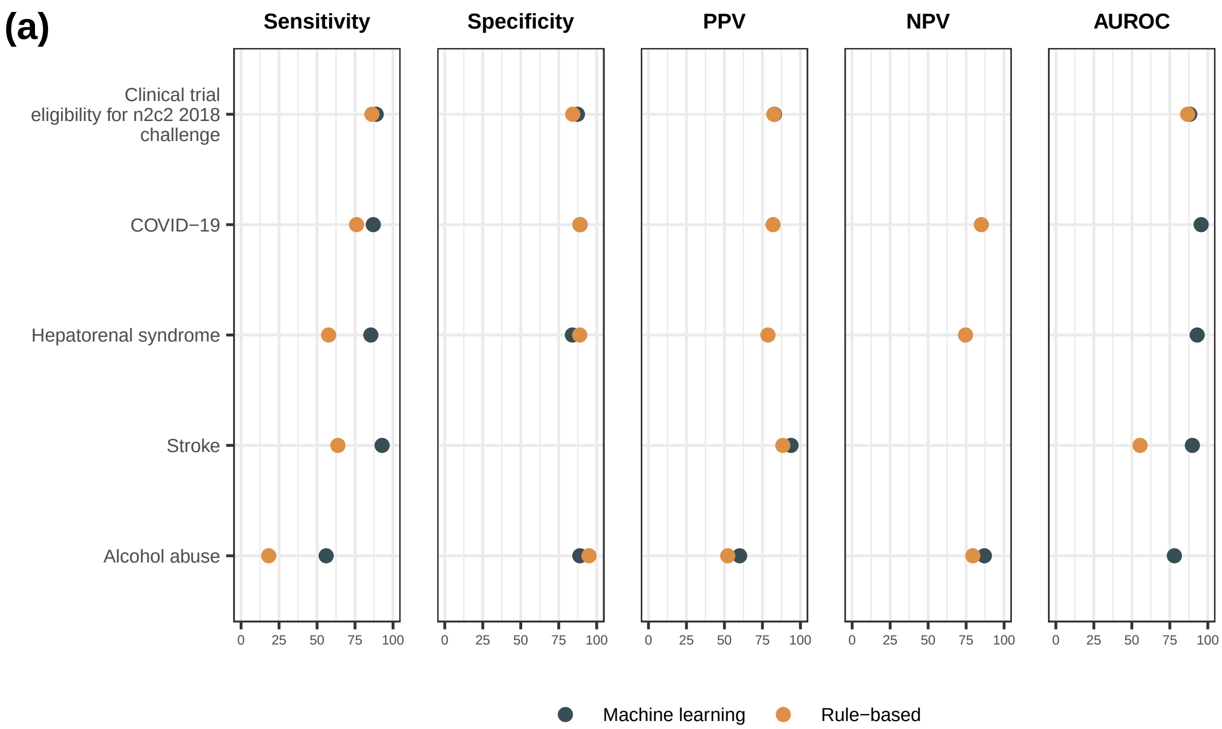
```
## [1] "There are 7 papers using multiple embedding training data"
```

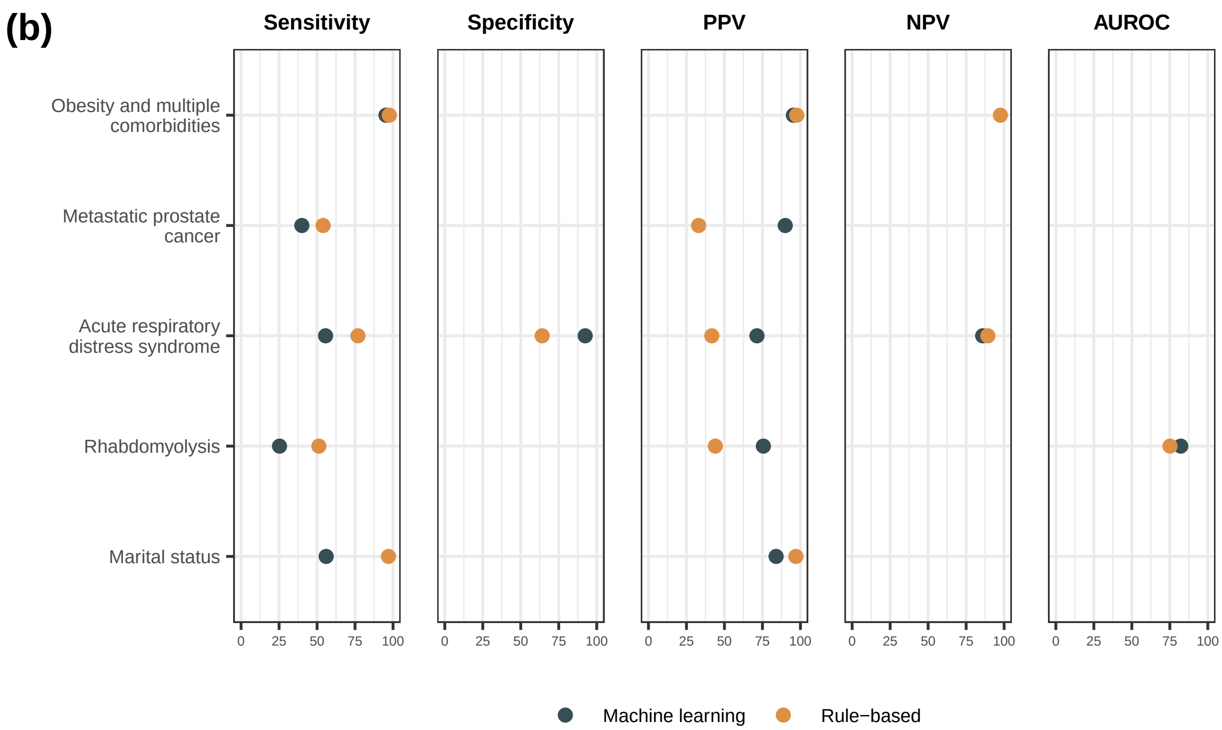
Embedding	Count
Word2vec	19
GloVe	6
BERT	5
RoBERTa	3
BioBERT	2
BioClinicalBERT	2
FastText	2
Not specified	2

```
## [1] "There are 11 papers using multiple embedding training methods"
```

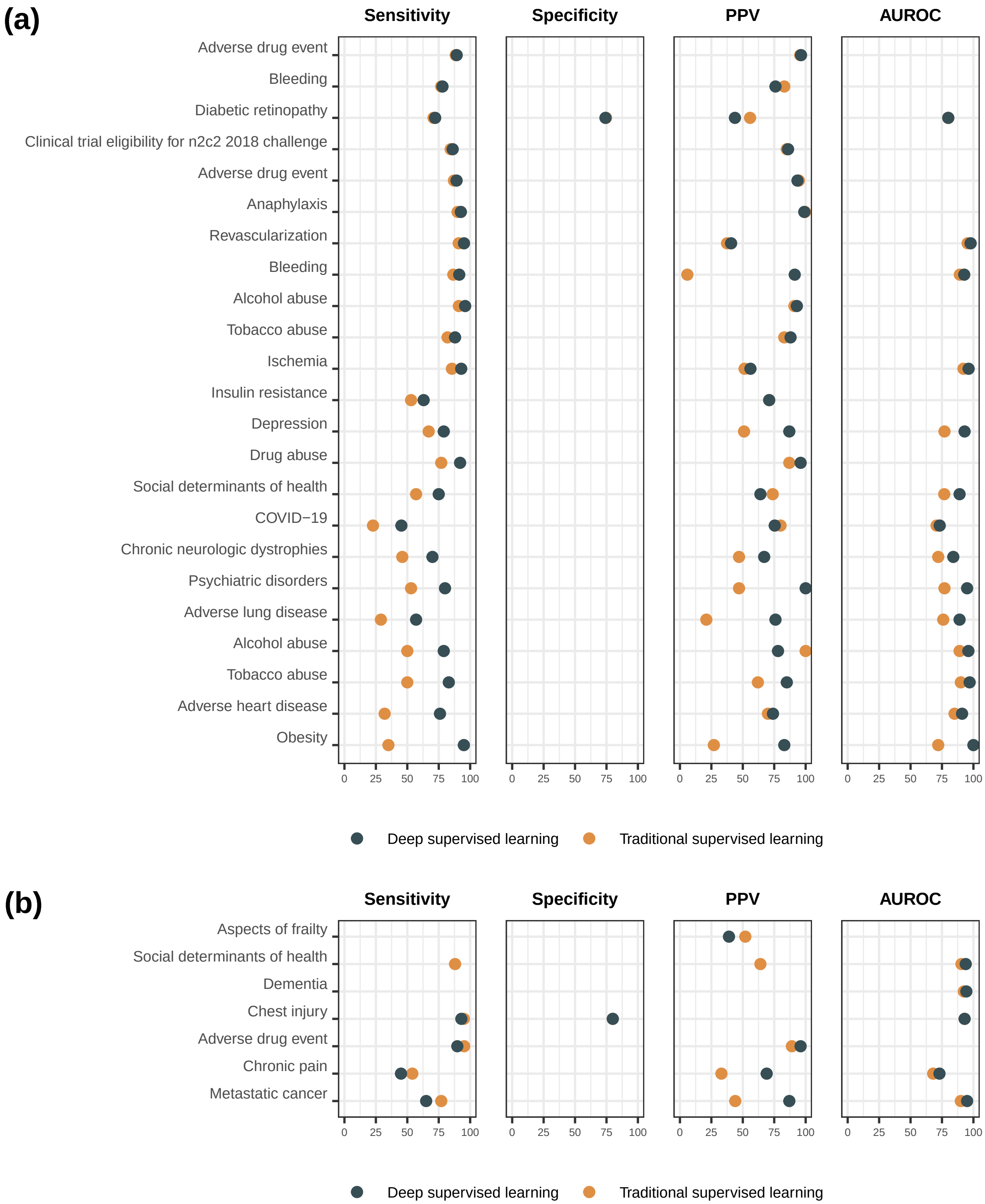
6 Validation and comparison

6.1 Traditonal supervised ML vs. rule-based





6.2 Deep supervised ML vs. traditional supervised ML



6.3 Weakly-supervised ML vs. rule-based algorithms

6.4 Weakly-supervised ML vs. traditional supervised ML

7 Model performance metric reporting

Model performance metrics	Supervised Deep learning	Supervised Traditional machine learning	Weakly-supervised Deep learning	Weakly-supervised Traditional machine learning	Reinforcement Deep learning	Semi-supervised Traditional machine learning	Count
Precision	26	23	0	8	0	4	61
Recall	26	23	1	7	0	2	59
AUROC	10	15	1	10	1	5	42
F-score	26	9	0	7	0	0	42
Specificity	7	11	1	1	0	0	20
Accuracy	5	8	1	4	0	0	18
NPV	1	7	0	5	0	2	15
AUPRC	3	2	0	2	1	0	8
Calibration plots	2	3	0	0	0	0	5
Log loss	1	1	0	0	0	1	3
Brier score	1	1	0	0	0	0	2
Hamming loss	2	0	0	0	0	0	2
Matthews Correlation Coefficient	1	1	0	0	0	0	2
Normalized discounted cumulative gain	1	1	0	0	0	0	2