Analysis of selected articles

Siyue Yang, Jessica Gronsbell

Last Updated: 10/17/2022

Contents

1 Machine learning (ML) methods **2**
1.1 ML paradigms 2
1.2 Traditional ML methods 2
1.3 Deep learning (DL) methods 3

2 Phenotypes **5**
2.1 Phenotypes considered across ML paradigms 5
2.2 Unstratified summary of phenotypes considered 6

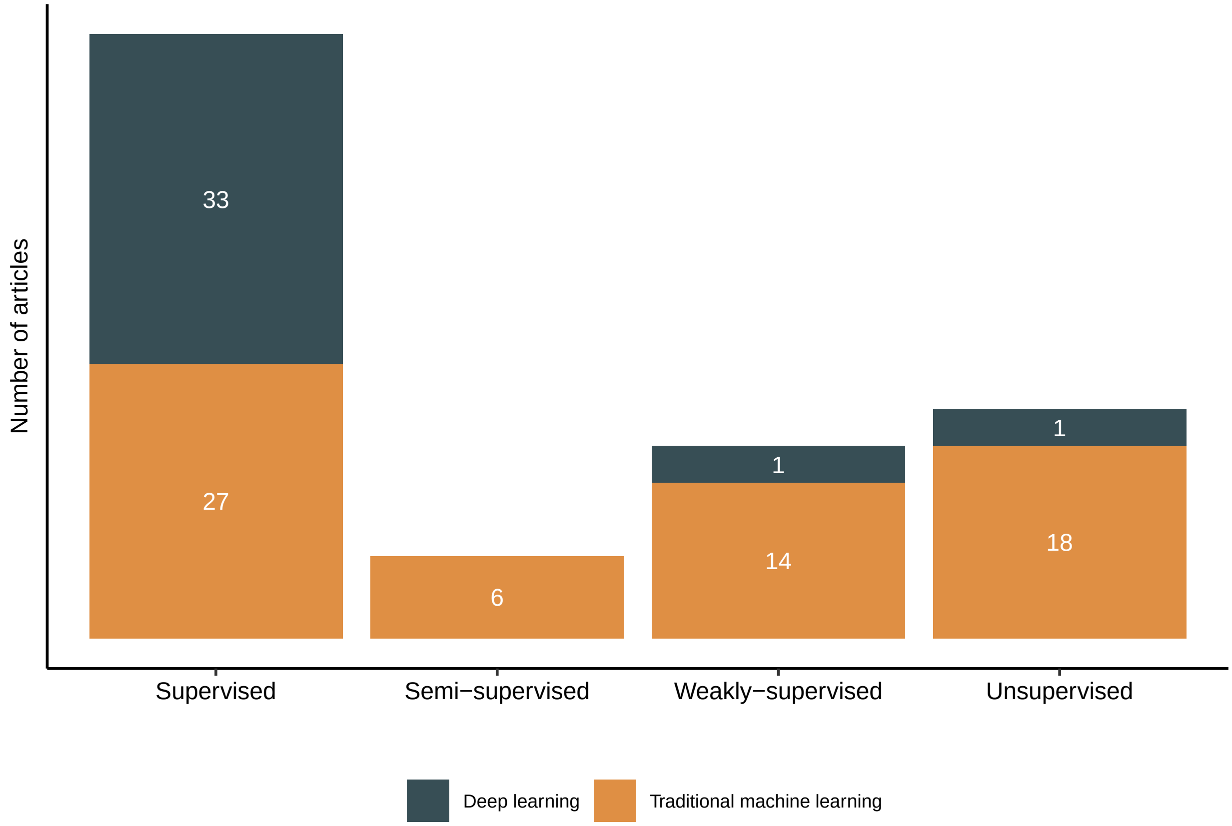
3 Data sources **7**
3.1 Use of structured and unstructured data 7
3.2 Structured and unstructured data types 8
3.3 Terminologies 9
3.4 Natural language processing (NLP) software 10
3.5 Embeddings 10
3.6 Openly-available data 11
3.7 Private data sources and demographics reporting 11
3.8 Institutions 12
3.9 Data sources summary different ML paradigms 12

4 Validation and comparison **14**
4.1 Traditonal supervised ML vs. rule-based 14
4.2 Deep supervised ML vs. traditional supervised ML 15

5 Model performance metric reporting **16**

1 Machine learning (ML) methods

1.1 ML paradigms



1.2 Traditional ML methods

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
Supervised	Decision trees	4
Supervised	XGBoost	4
Supervised	Naive Bayes	3
Weakly-supervised	PheNorm	3
Weakly-supervised	MAP	2
Weakly-supervised	Random forest	2
Unsupervised	LDA	5
Unsupervised	K-means	4
Unsupervised	UPGMA Hierarchical clustering	2

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

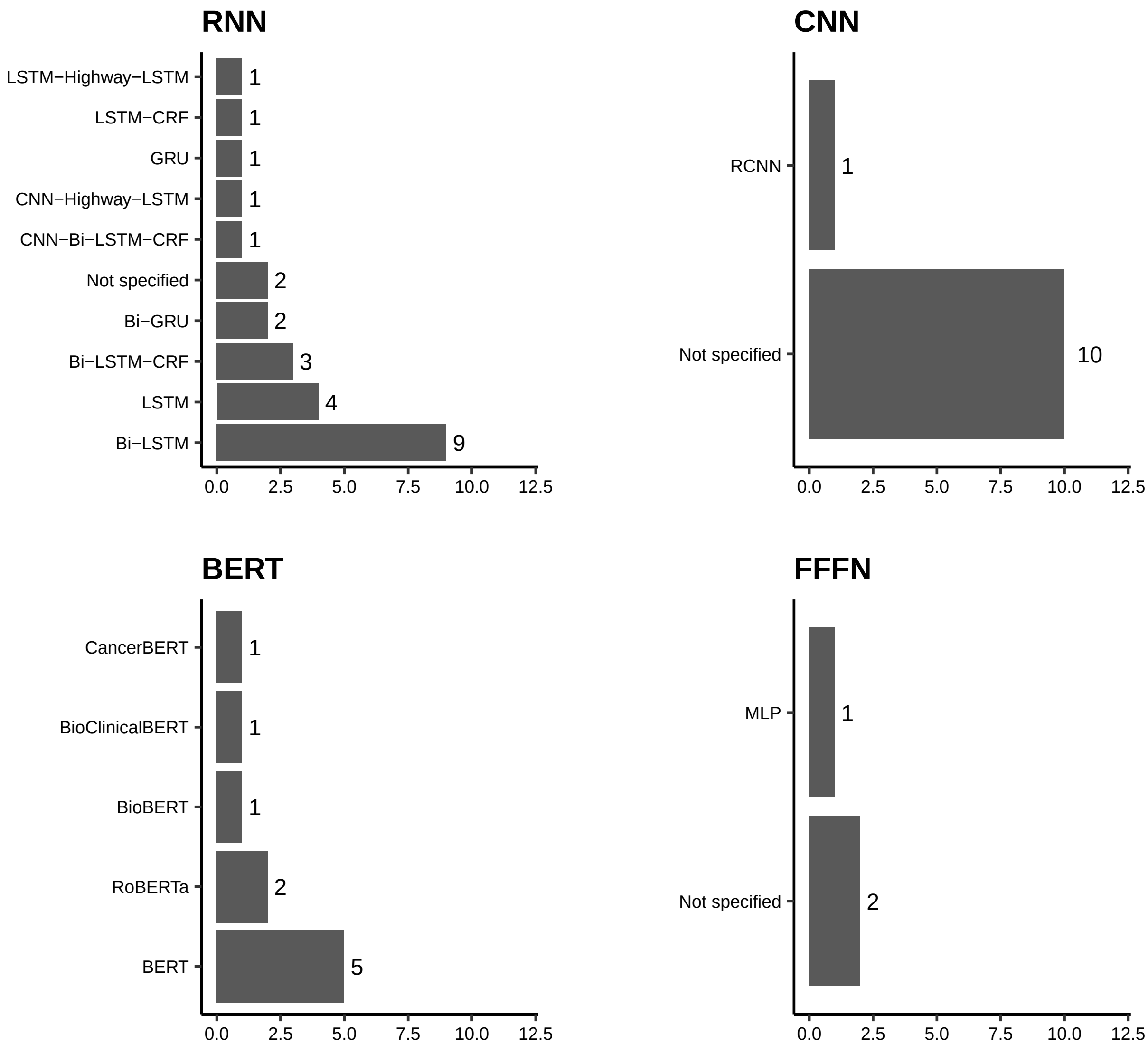
1.3 Deep learning (DL) methods

Table 2: Common deep learning methods (Count > 1)

DL method	ML	Count
BERT	Supervised	7
CNN	Supervised	11
FFNN	Supervised	3
RNN	Supervised	19

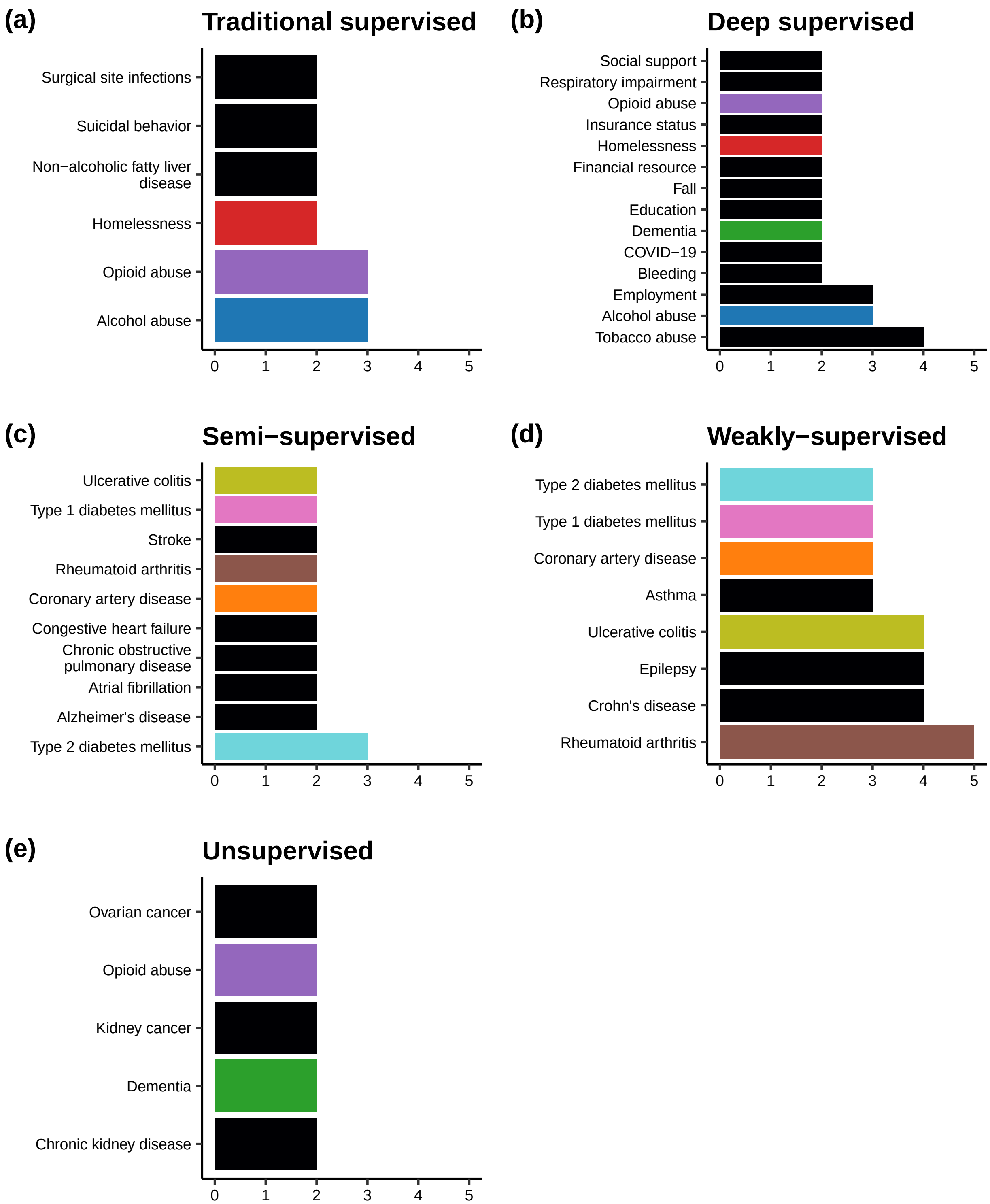
```
## [1] "There are 5 papers using multiple deep learning methods"
```

1.3.1 Neural network variants

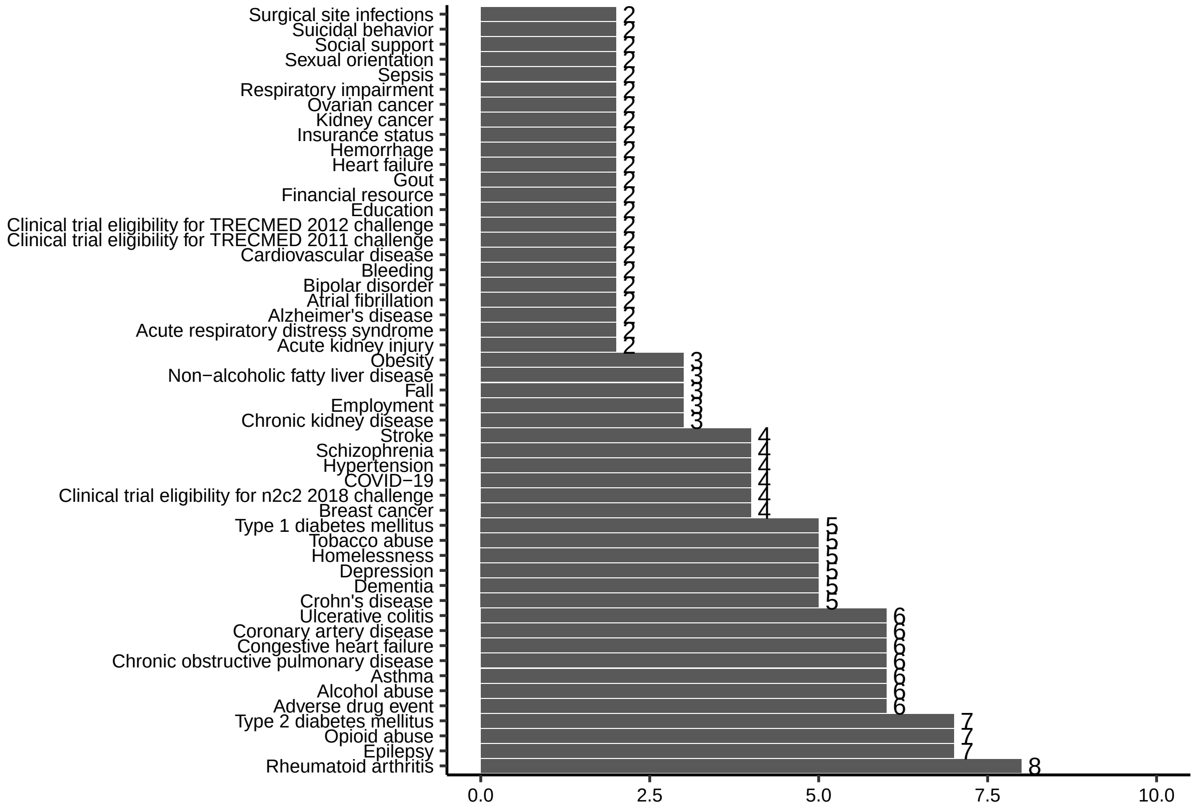


2 Phenotypes

2.1 Phenotypes considered across ML paradigms



2.2 Unstratified summary of phenotypes considered



3 Data sources

3.1 Use of structured and unstructured data

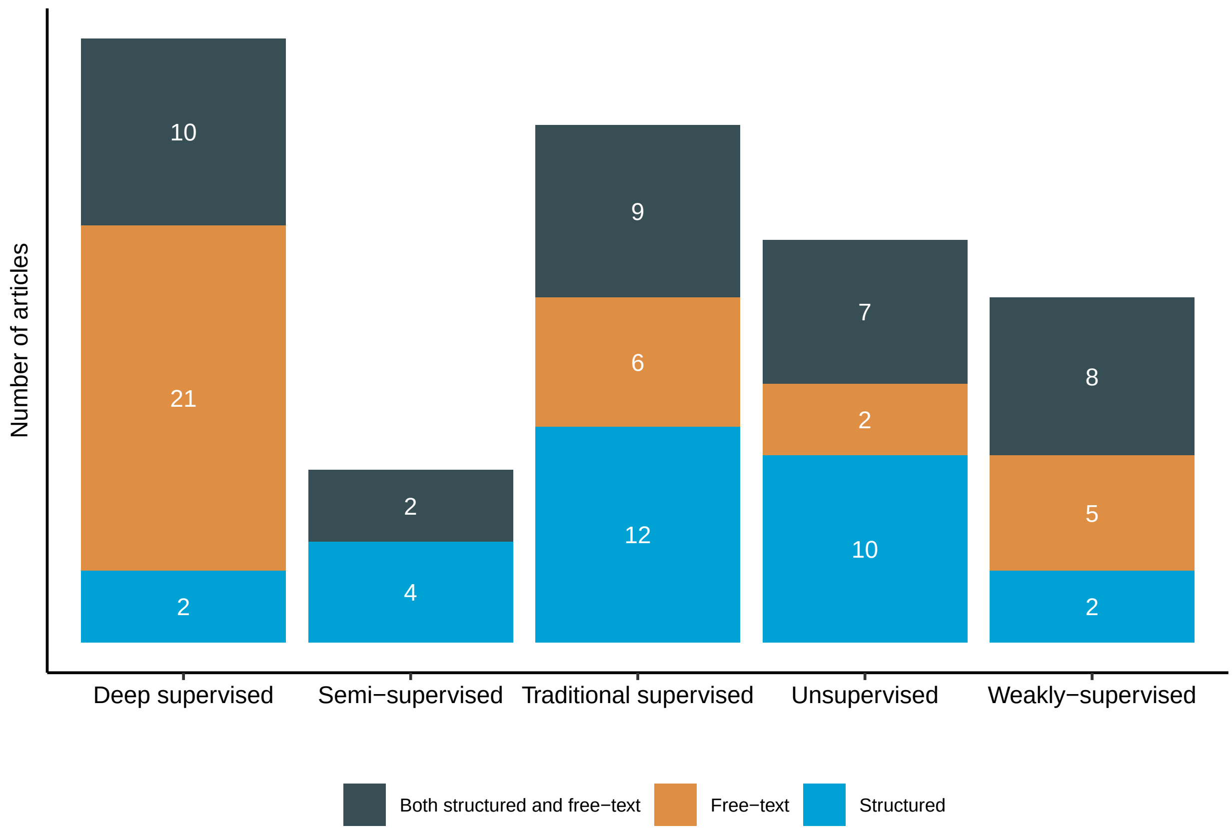


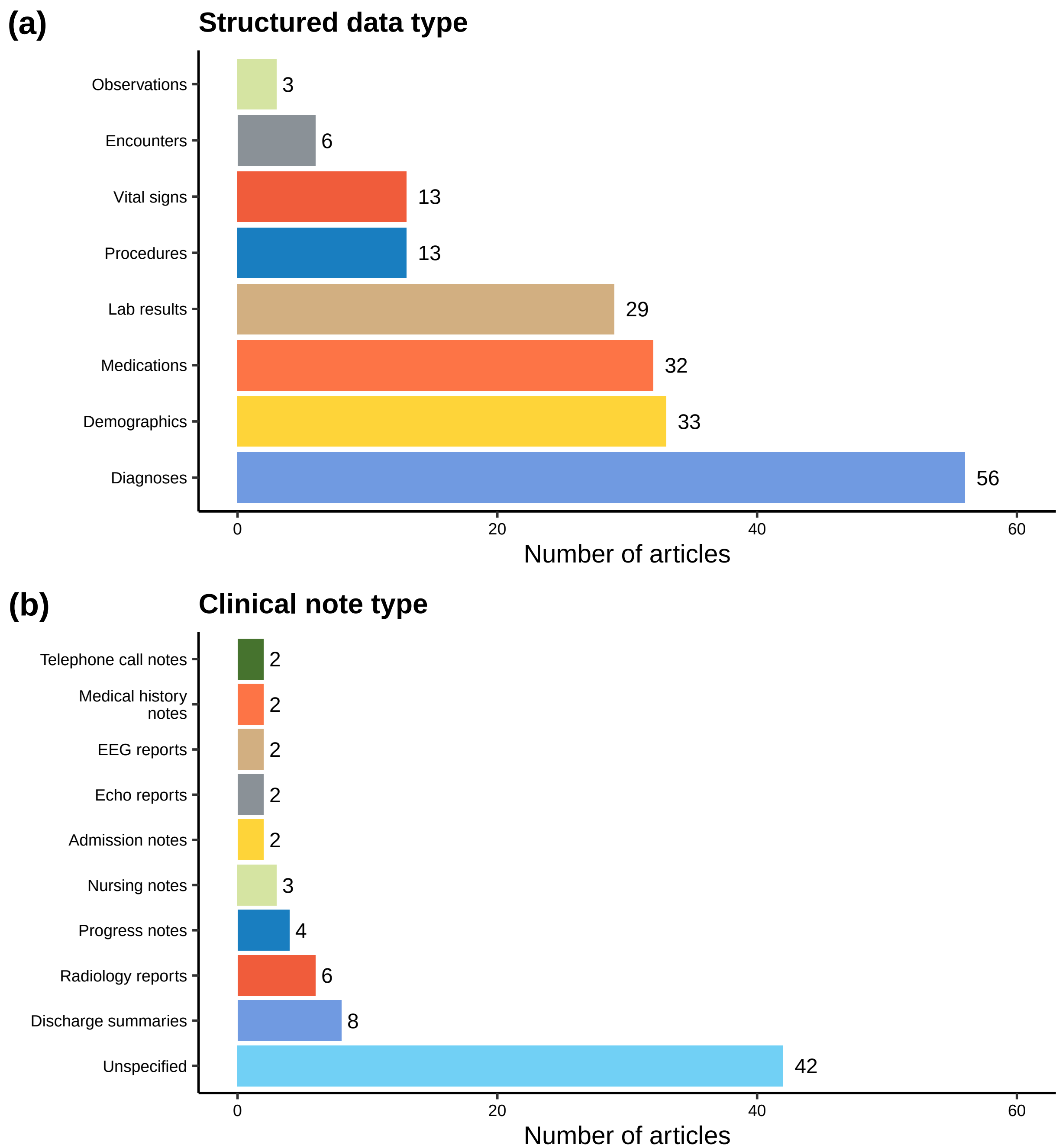
Table 3: Use of structured and unstructured data

Data	Count
Both structured and free-text	36
Free-text	34
Structured	30

3.2 Structured and unstructured data types

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

## [1] "There are 0 papers using multiple unstructured data types"
```



Terminology unnested	Supervised Traditional machine learning	Unsupervised Traditional machine learning	Supervised Deep learning	Weakly-supervised Traditional machine learning	Semi-supervised Traditional machine learning	Count
ICD-9	18	8	7	5	4	42
UMLS	11	3	8	8	1	31
ICD-10	11	2	4	1	3	21
SNOMED-CT	2	3	4	3	0	12
RxNorm	3	1	2	2	1	9
CPT	2	0	3	2	0	7
Phecode	0	2	0	3	2	7
ICD	0	1	0	4	0	5
LOINC	3	0	0	1	0	4
ATC (Anatomical therapeutic chemical)	2	0	0	0	0	2
NDC (National drug codes)	2	0	0	0	0	2

3.3 Terminologies

```
## [1] "There are 43 papers using multiple terminologies"
```

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

3.4 Natural language processing (NLP) software

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

3.5 Embeddings

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"
```

Data source	Supervised Deep learning	Supervised Traditional machine learning	Weakly-supervised Deep learning	Weakly-supervised Traditional machine learning	Unsupervised Traditional machine learning	Count
MIMIC-III database	9	1	1	1	3	15
MTSamples database	1	0	0	0	0	1

3.6 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
2008 i2b2	1	0	1
2012 physionet Challenge	0	1	1

3.7 Private data sources and demographics reporting

```
## [1] "71 articles did not use openly available data"
```

```
## [1] "Among these 71 articles, 38 articles considered temporal phenotypes"
```

3.8 Institutions

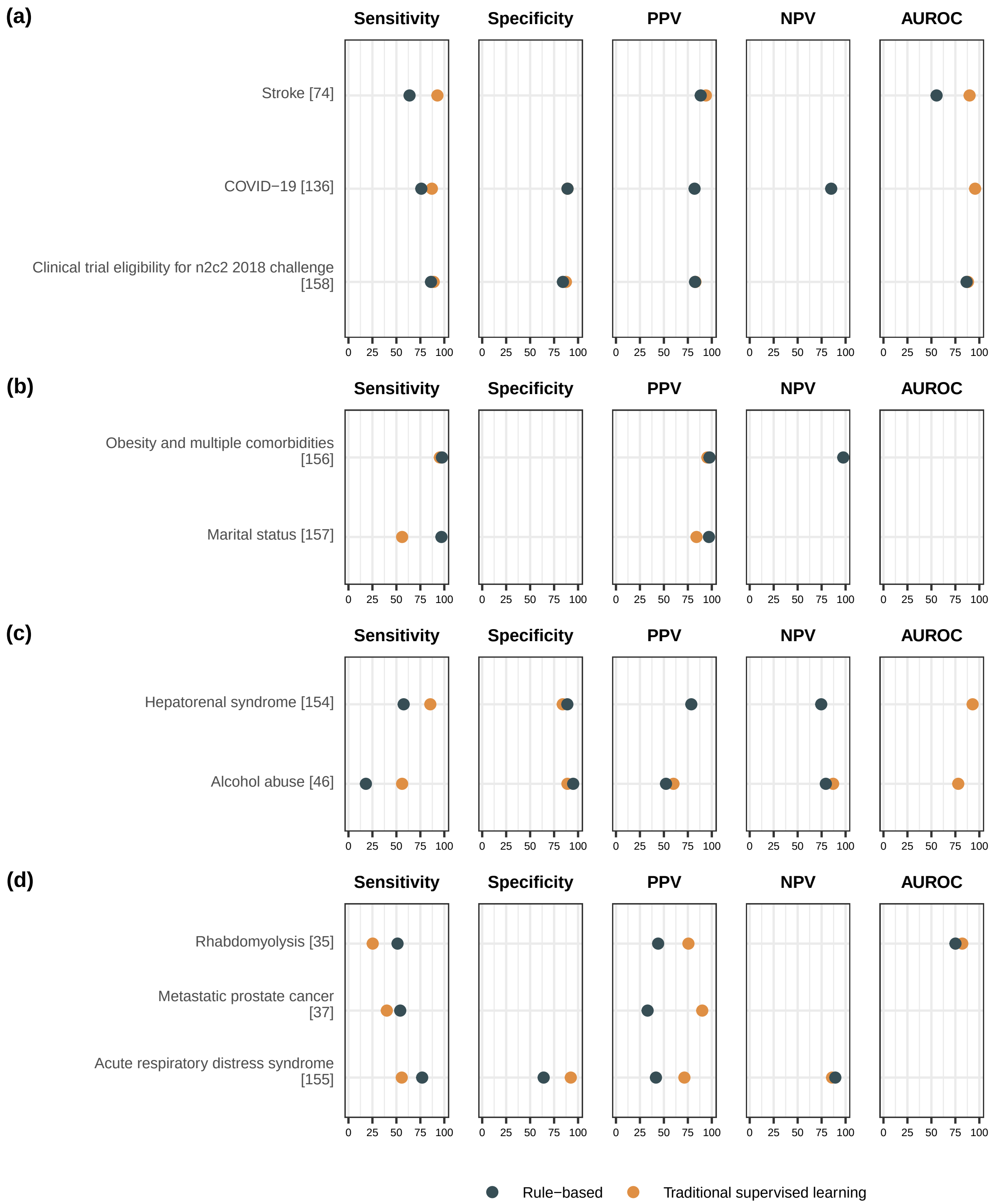
Country	Count
US	94
France	2
Cancada	1
China	1
Germany	1
Israel	1
Italy	1
Korean	1
Netherland	1
Singapore	1
Spain	1

3.9 Data sources summary different ML paradigms

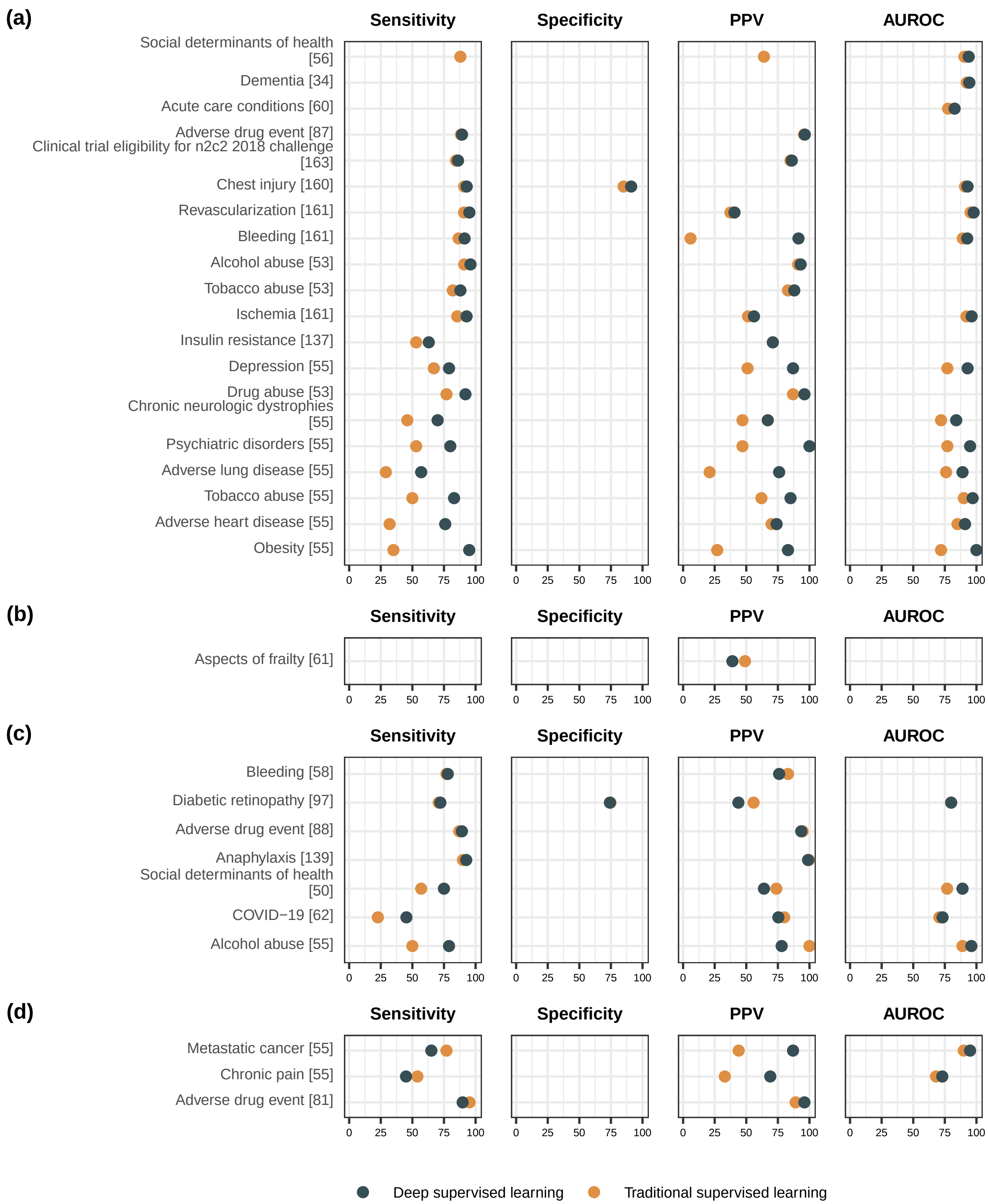
	Total number of papers	Used free-text	Used NLP software	Used competition data	Used multisite data	Used open data	Used private single-site data	Compared to rule-based algorithms	Comapred to traditional ML	Reported patient demographic	Released open code
TSL	27	15	14	3	1	1	22	10	0	13	4
DSL	33	31	18	11	1	9	12	2	20	5	9
SSL	6	2	1	0	0	0	6	1	0	3	0
WSL	15	13	10	0	3	2	10	8	1	4	3
USL	19	9	4	0	3	3	13	0	0	13	4
Total	100	70	47	14	8	15	63	21	21	38	20

4 Validation and comparison

4.1 Traditonal supervised ML vs. rule-based



4.2 Deep supervised ML vs. traditional supervised ML



Model performance metrics	Supervised Deep learning	Supervised Traditional machine learning	Weakly-supervised Deep learning	Weakly-supervised Traditional machine learning	Semi-supervised Traditional machine learning	Count
Precision	26	23	0	8	4	61
Recall	25	23	1	7	2	58
AUROC	11	15	1	10	5	42
F-score	26	9	0	7	0	42
Specificity	6	11	1	1	0	19
Accuracy	4	8	1	4	0	17
NPV	1	7	0	5	2	15
AUPRC	4	2	0	2	0	8
Calibration plots	2	3	0	0	0	5
Log loss	1	1	0	0	1	3
Brier score	1	1	0	0	0	2
Hamming loss	2	0	0	0	0	2
Matthews Correlation Coefficient	1	1	0	0	0	2
Normalized discounted cumulative gain	1	1	0	0	0	2

5 Model performance metric reporting