Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for EHR Analysis

2018

Benjamin Shickel, Patrick J Tighe, Azra Bihorac, Parisa Rashidi
University of Florida

Presenter: Derrick Blakely
@ https://qdata.github.io/deep2Read/

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Background - Electronic Health Records (EHRs)

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data

Background

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data
- Hospital admission and discharge data (datetime objects)
- Lab tests/results
- Radiological images

Background

- Huge increases in the numbers of EHRs in the US in the last 10 years
- Heterogeneous data

- Hospital admission and discharge data (datetime objects)
- Lab tests/results
- Radiological images
- Genomic data
- ICD codes
- Clinical notes (free text)

EHRs

- Primary use: bookkeeping, hospital administration

EHRs

- Primary use: bookkeeping, hospital administration
- Secondary uses:
  - Medical concept extraction
  - Patient trajectory modeling
  - Disease inference
  - Clinical decision support systems
  - Deidentification
  - Phenotyping

EHR Analysis

- Traditional: logistic regression, random forests, SVM

EHR Analysis

- Traditional: logistic regression, random forests, SVM
- Recent: MLP, Autoencoder, RBM, Deep Belief Nets, CNN, RNN, GRU, and LSTM

EHR Analysis

- Traditional: logistic regression, random forests, SVM
- Recent: MLP, Autoencoder, RBM, Deep Belief Nets, CNN, RNN, GRU, and LSTM
- Most “Deep EHR” papers published in last 3 years
  - Several hundred total

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Motivation

- Catalog advances
- High-level overview of what’s been going on in EHR analysis in the last few years
- Future directions

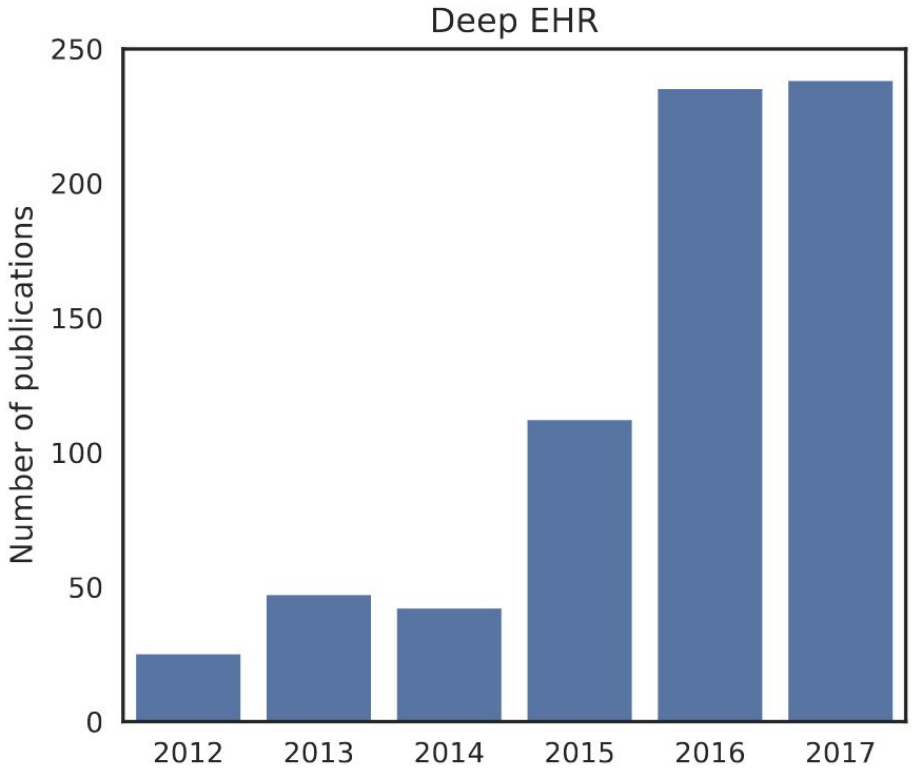
Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

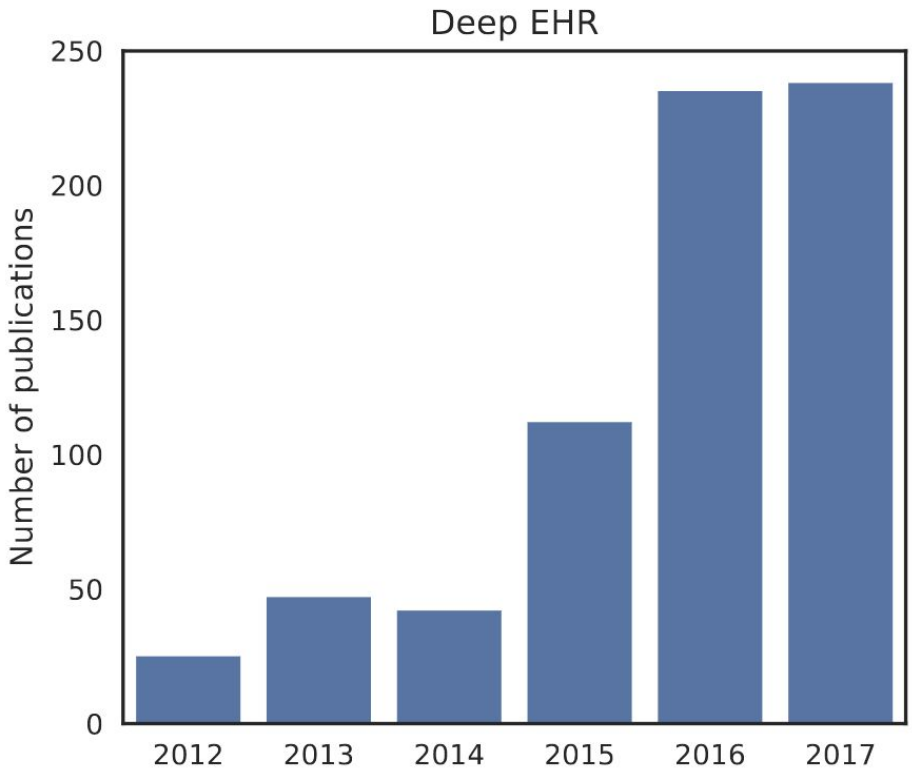
Roadmap

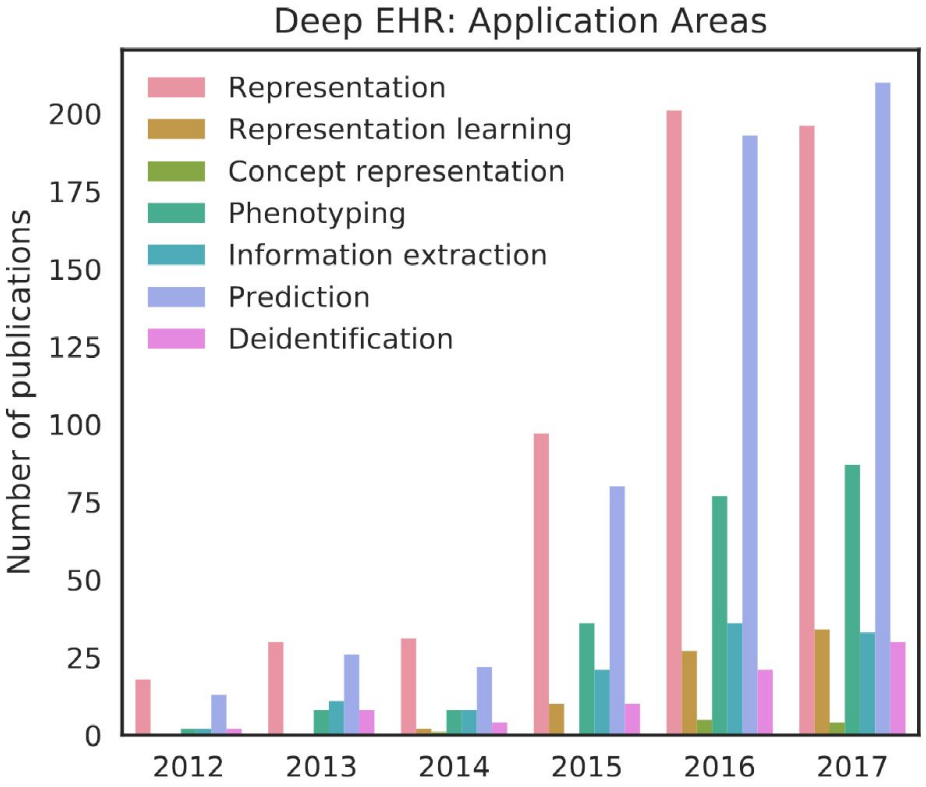
1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Deep EHR Overview

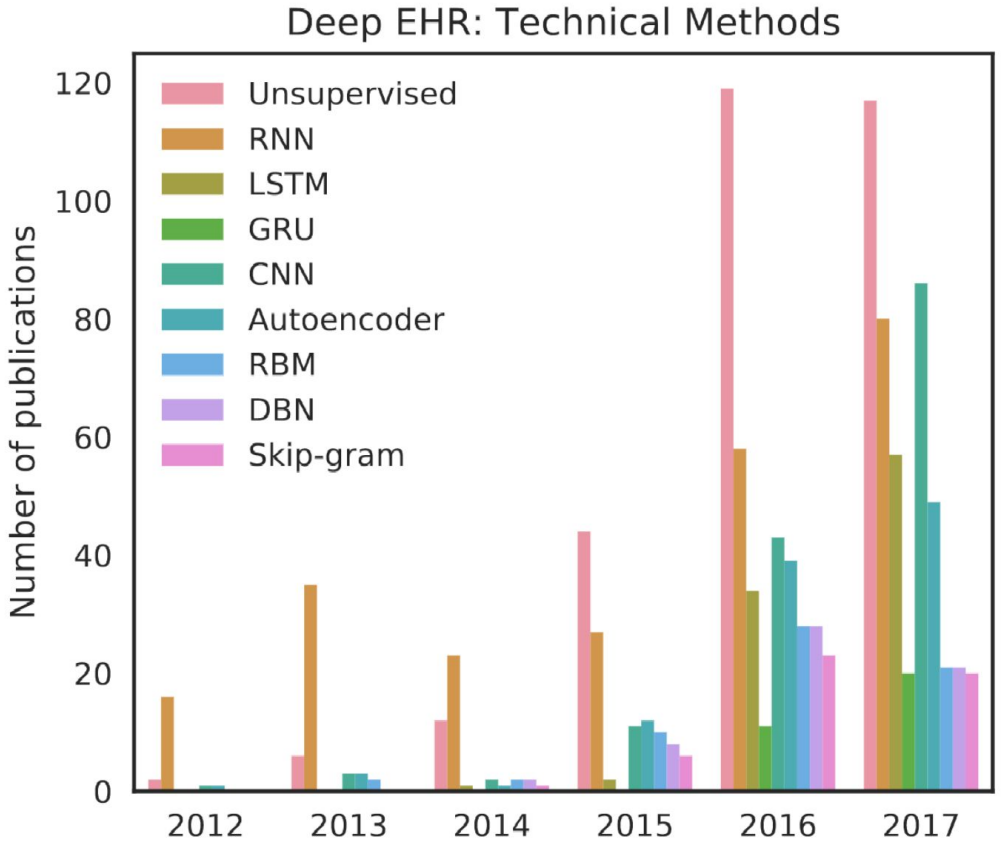


Deep EHR Overview





Deep EHR Overview



Deep EHR Overview

Task	Subtasks	Input Data	Models
Information Extraction	(1) Single Concept Extraction (2) Temporal Event Extraction (3) Relation Extraction (4) Abbreviation Expansion	Clinical Notes	LSTM, Bi-LSTM, GRU, CNN RNN + Word Embedding AE Custom Word Embedding
Representation Learning	(1) Concept Representation (2) Patient Representation	Medical Codes	RBM, Skip-gram, AE, LSTM RBM, Skip-gram, GRU, CNN, AE
Outcome Prediction	(1) Static Prediction (2) Temporal Prediction	Mixed	AE, LSTM, RBM, DBN LSTM
Phenotyping	(1) New Phenotype Discovery (2) Improving Existing Definitions	Mixed	AE, LSTM, RBM, DBN LSTM
De-identification	Clinical text de-identification	Clinical Notes	Bi-LSTM, RNN + Word Embedding

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Roadmap

1. Background
2. Motivation
3. Survey of Recent Advances in “Deep EHR”
4. Future Directions

Future Directions: Representations

- Representations in light large amount of heterogeneity

Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes

Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings

Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings
- Clinical text is under-utilized

Future Directions: Representations

- Representations in light large amount of heterogeneity
- Heavy focus on clinical codes
- Many things are not incorporated into representations/embeddings
- Clinical text is under-utilized
- “Holy grail”: unified representation

Future Directions: Benchmarks

- Lack of universal benchmarks
- Difficult reproducibility
- Everyone claims “state-of-the-art performance”
- Proprietary data sets
- Hyperparameters can make or break an algorithm

Future Directions: Interpretability

- Models need to be transparent and trustworthy
- Explored so far: maximum activation, clustering illustrations, word clouds, heat maps, “Mimic learning”