**Opportunities and Challenges in Developing Deep Learning Models Using Electronic Health Records Data: a Systematic Review**

**Supplementary File**

Task Category:

A. disease detection or classification,

B. sequential prediction of clinical events,

C. concept embedding

D. data imputation and generation,

E. EHR data privacy.

Venue Type:

I. medical journals

II. informatics journal or conferences

III. computer science

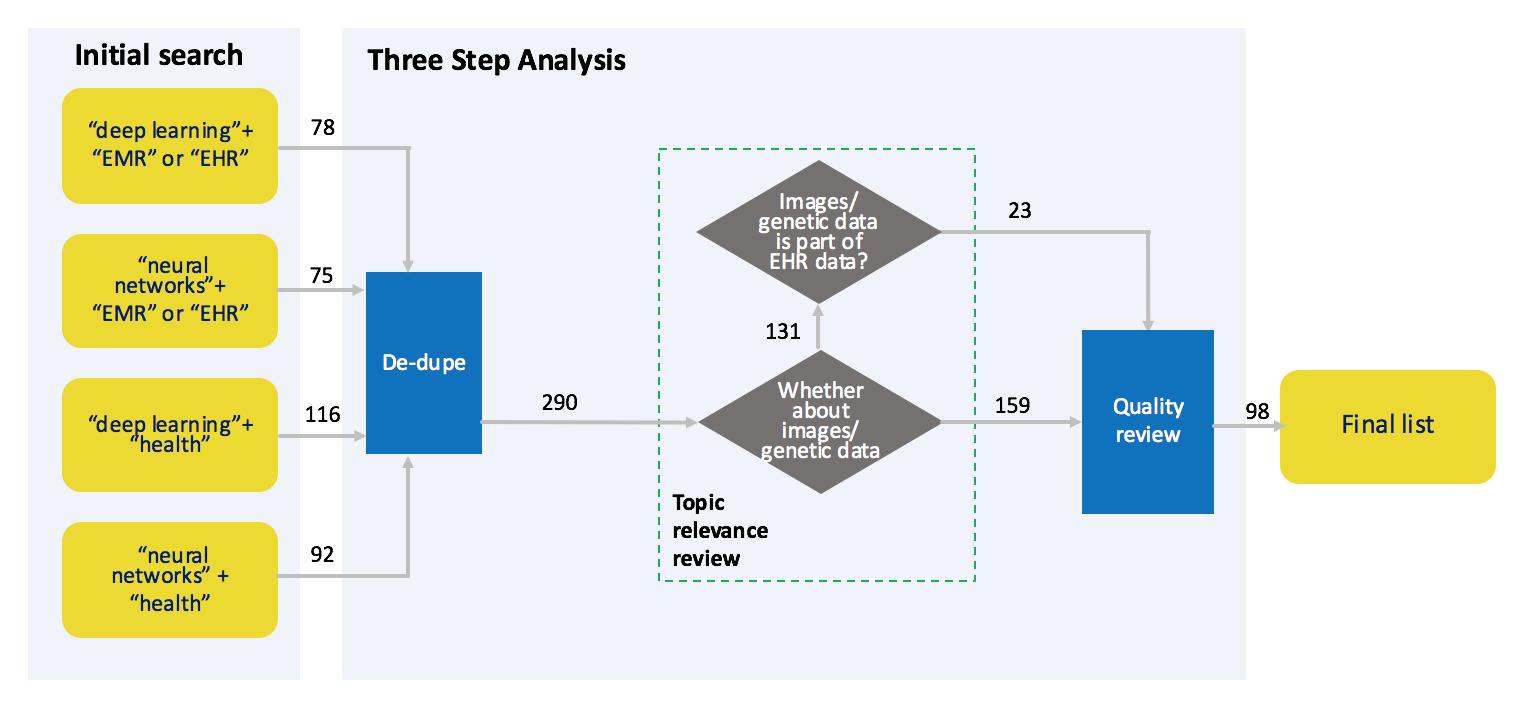
Table S1: Characteristics of Studies Included in the Review

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Venue | | | Data | | | Task and Model | | | | |
| Paper | Venue | Venue Type | Year | Data | Patient | Record | Task | Task Category | Time Horizon | Model Type | Performance |
| [[12]](https://paperpile.com/c/26vM9E/m1Zpw) | SDM | III | 2016 | Proprietary EHR | 319,650 | NA | CHF and COPD prediction | A | 4 yrs | Temporal fused CNN | AUC ~0.7675 (CHF prediction); ~0.7388 (COPD prediction) |
| [[49]](https://paperpile.com/c/26vM9E/VWT57) | JBI | III | 2017 | Australian proprietary | 7000 diabetes; 6100 mental health | 53,000 & 52,000 | readmission prediction | B | 2002-2013 | LSTM | AUC: 0.718 & 0.726 |
| [[30]](https://paperpile.com/c/26vM9E/Djp8) | NSR | II | 2016 | Mount Sinai data warehouse. | 700,000 | NA | Unsupervised embedding & prediction | C | 2014 | Stacked denoising autoencoders | AUC 0.773 |
| [[56]](https://paperpile.com/c/26vM9E/ADfYq) | PAKDD | III | 2016 | Australian proprietary | 7191 | 53,208 admissions | readmission prediction | B | 2002-2013 | LSTM | F1: 0.79 |
| [[31]](https://paperpile.com/c/26vM9E/UOdXC) | KDD | III | 2016 | CHOA & CMS | 550,339 & 831,210 | 3,359,240;  5,464,950 | Embedding & eval with prediction | C | 8 yrs | Skip gram | AUC: 0.760 (code) & 0.915 (class) |
| [[26]](https://paperpile.com/c/26vM9E/7NgMs) | MLHC | III | 2016 | Sutter Health EHR & MIMIC II | 263,706 & 2695 | 14,366,030 | Multilabel sequential  prediction | B | 8 yrs & 7 yrs | RNN+skip gram | Recall 79.58% |
| [[43]](https://paperpile.com/c/26vM9E/uHurb) | NIPS | III | 2016 | Sutter Health EHR | 263,683 | 14,366,030 visits | Heart failure prediction | B | 8 yrs | RNN + attention model | AUC: 0.8705 |
| [[44]](https://paperpile.com/c/26vM9E/os23r) | KDD | III | 2017 | Sutter PAMF & MIMIC III & Sutter HF | 258,555; 7,499;  30,727 | 13,920,759;  19,911;  572,551 | Sequential & HF onset prediction | B | 10 yrs; 11 yrs | Ontology regularized GRU | AUC ~0.845  (HF onset prediction) |
| [[36]](https://paperpile.com/c/26vM9E/7JkL3) | NIPS ML4H | III | 2017 | New York Presbyterian/Columbia University  EHR | 485,306 | 473.6 mil lab tests | predicting drug-Induced  laboratory test | D | 2000-2013 | GANs | MSE ~0.11-0.15 |
| [[23]](https://paperpile.com/c/26vM9E/vfOZT) | JAMIA | II | 2017 | Sutter Health EHR | 32787 | 58,652,000 codes | Predict HF onset | B | 2000-2013 | RNN | AUC: 0.883 |
| [[98]](https://paperpile.com/c/26vM9E/0tpNC) | JBI | II | 2015 | Australian proprietary | 7,578 | 17,566 assessments | concept embedding & eval with suicide risk prediction | C | 01/2009 - 03/2012. | RBM | F1: 0.212 (moderate risk) & 0.359 (high risk) |
| [[58]](https://paperpile.com/c/26vM9E/qmPSg) | ICHI | II | 2016 | Charite Hospital in Berlin | 4,000 | NA | Predict endpoint of kidney transplant | B | 30 years | RNN | AUC ~0.833 |
| [[85]](https://paperpile.com/c/26vM9E/KB34) | KDD | III | 2015 | PhysioNet Challenge 2012 & ICU data | na | 3940 episodes;  8500 episodes | concept embedding  for vital signs | C | na | DAE | AUC ~0.74 |
| [[10]](https://paperpile.com/c/26vM9E/C1B2) | JBI | II | 2016 | Open access ALS Clinical Trials (PRO-ACT) | 10723 | NA | concept embedding | C | na | denoising autoencoders | Mean AUC ~0.692 |
| [[86]](https://paperpile.com/c/26vM9E/YQEj) | PLoS One | II | 2013 | Proprietary EHR | 4368 | NA | concept embedding & eval with classification | C | na | autoencoders | AUC 0.972 |
| [[37]](https://paperpile.com/c/26vM9E/bUprs) | JAMIA | II | 2017 | i2b2 2014 dataset & MIMIC III | 46520 (MIMIC III) | 984,723 tokens; 2,945,228 tokens | Data de-identification | E | 2004; 2000-2011 | bi-LSTM | F1 97.877 (i2b2); 99.108 (MIMIC III) |
| [[94]](https://paperpile.com/c/26vM9E/uuwSi) | AMIA | II | 2016 | Pediatric ICU@CHLA | 398 | NA | Mimic learning for interpretation; ICU outcome prediction | B | na | GBTmimic | AUC 0.7898 (Mortality prediction);  0.7889 (Ventilator free days prediction) |
| [[117]](https://paperpile.com/c/26vM9E/bIZtv) | JMLR | III | 2016 | Type 2 Diabetes data@Duke University Health System | 16756 | NA | Multi-modal Comp. Phenotyping | B | 2007-2011 | Deep Poisson Factor Analysis | AUC 0.784 |
| [[50]](https://paperpile.com/c/26vM9E/dBSm9) | NAACL-HLT | III | 2016 | Notes of patients  diagnosed w. hematological malignancy. | 780 notes | NA | Sequential labeling of clinical notes (for ADEs) | A | na | RNN | F1  0.8031 |
| [[51]](https://paperpile.com/c/26vM9E/qFU1A) | EMNLP | III | 2016 | EHR or cancer patients | 1154 records | NA | Label clinical events (classification) from clinical notes | A | na | LSTM | F1 0.821 |
| [[20]](https://paperpile.com/c/26vM9E/dHdmC) | arxiv | III | 2017 | MIMIC II & III | 20,533 notes 49,857 notes | NA | Multi-label clinical notes classification | A | 01-08; 00-11 | CNN+Hierarchical attention GRU | Micro F1: 53.86% 55.86% |
| [[45]](https://paperpile.com/c/26vM9E/dKEqV) | arxiv | III | 2016 | MIMIC III Clinical notes | 40000 | NA | Multilabel classification to tag patient notes with ICD-9 codes | A | 2001-2011 | LSTM | F1 0.715 |
| [[46]](https://paperpile.com/c/26vM9E/Q0xco) | ICLR | III | 2016 | Pediatric ICU | 10401 | NA | Diagnostic classification using multivariate ICU time series | A | NA | LSTM | AUC 0.8643 |
| [[11]](https://paperpile.com/c/26vM9E/QPiJO) | KDD | III | 2017 | Parkinson’s Progression Markers Initiative & synthetic EHR | 654 (PPMI);  100,000 (synthetic EHR) | NA | concept embedding | C | NA | LSTM | AUC 0.91 (synthetic EHR); MSE 0.50 (PPMI) |
| [[97]](https://paperpile.com/c/26vM9E/ZJSvA) | EMBC | II | 2016 | MIMIC II | 15,647 patients | NA | Predict patient mortality at discharge | B | 2001-2008 | DBN | Accuracy 0.852 |
| [[24]](https://paperpile.com/c/26vM9E/0m36E) | JBI | II | 2015 | New Zealand National Minimum  Dataset | 1,328,384 | 3,295,775 | Readmission prediction | B | 2006-2012 | DNN | AUC 0.638~0.734 |
| [[52]](https://paperpile.com/c/26vM9E/3DRbw) | NIPS ML4H | III | 2017 | Nokia fitness data | 15000 | 18036 sequences | Predict whether users will achieve their weight objectives | B | 2-month | LSTM variant | AUC 0.8807 |
| [[95]](https://paperpile.com/c/26vM9E/ehzzA) | BIBM | II | 2014 |  |  |  | Diagnostic classification | A |  | DNN |  |
| [[81]](https://paperpile.com/c/26vM9E/IMM6) | BCB | II | 2016 | Osteoporotic Fractures (SOF) dataset | 9704 | NA | Disease prediction | A | NA | RBM/autoencoders | AUC 0.864 |
| [[127]](https://paperpile.com/c/26vM9E/ChMt3) | ICML | III | 2016 | semi-simulated dataset based on  the Infant Health and Development Program (IHDP) | 747 | NA | causal  inference (Counterfactual Inference) |  | NA | DNN variant | MSE ~0.02 |
| [[82]](https://paperpile.com/c/26vM9E/oMBC) | BCB | II | 2017 | CHB-MIT Scalp EEG Database | 23 | NA | Seizure detection | A | NA | stacked sparse denoising  autoencoders | AUC 0.9833 |
| [[34]](https://paperpile.com/c/26vM9E/Th7Os) | ICDM | III | 2017 | Proprietary EHR | 218, 680 | 14, 969, 489 | EHR generation | D | 2011-2015 | GANs |  |
| [[71]](https://paperpile.com/c/26vM9E/j16Bl) | Computing in Cardiology Conference | II | 2016 | PCG recordings | 3240 samples | NA | Diagnostic classification | A | NA | CNN+ensemble | AUC 0.85 |
| [[21]](https://paperpile.com/c/26vM9E/6KdB5) | Advances in Big Data | III | 2016 | Cancer pathology reports | NA | NA | Automated pathology report annotation | A | NA | DNN | Macro F-score 0.998 (primary site); 0.948 (laterality) |
| [[32]](https://paperpile.com/c/26vM9E/RbbUw) | arxiv | III | 2017 | MIMIC III | NA | 1,610 discharge summaries | concept embedding | C | 2011-2015 | CNN | F1  0.76 |
| [[57]](https://paperpile.com/c/26vM9E/oa3Iw) | JBI | II | 2017 | Australian proprietary | 7191 | 53,208 admissions | readmission prediction | B | 2002-2013 | LSTM | F-score  0.79 |
| [[115]](https://paperpile.com/c/26vM9E/KFV0n) | arxiv | III | 2017 | IHDP data | 747 | NA | Causal inference |  | NA | DNN variant | MSE  2.05 |
| [[59]](https://paperpile.com/c/26vM9E/Mpz3g) | arxiv | III | 2017 | MIMIC III | 34,148 | 34,148 ICU stays | multimodal  Prediction for both onset and weaning of interventions | B | 2011-2015 | LSTM+CNN | AUC ~0.9 |
| [[74]](https://paperpile.com/c/26vM9E/gQDOz) | NeuroImage: Clinical | I | 2017 | Data from EEG/fMRI study | 445 | NA | interictal epileptic discharges detection based on time series | B | 2006-2016 | CNN | Concordance  0.769 |
| [[13]](https://paperpile.com/c/26vM9E/IVSIZ) | Computers in Biology and Medicine | II | 2017 | MIMIC III | NA | NA | Sepsis detection | A | NA | DNN+LSTM | AUC  0.887 |
| [[15]](https://paperpile.com/c/26vM9E/LwWP0) | Computers in Biology and Medicine | II | 2017 | EEG | NA | NA | Seizure detection | A | NA | CNN | ACC 0.8867 |
| [[67]](https://paperpile.com/c/26vM9E/PHnGt) | SDM | III | 2017 | PPMI | 617 | 15636 | concept embedding | C | NA | LSTM | Micro F1: 0.7667 |
| [[47]](https://paperpile.com/c/26vM9E/BdDti) | KDD | III | 2017 | Medicaid claims | 147, 810 | 1, 055, 011 visits | prediction | B | 2011 | RNN | acc@k ~0.8 |
| [[70]](https://paperpile.com/c/26vM9E/d6Nk2) | ML4H | III | 2016 | Pediatric ICU | 10, 401 episodes | NA | Data imputation | D | NA | RNN | Micro AUC 0.873 |
| [[48]](https://paperpile.com/c/26vM9E/3putj) | AMIA TBI | II | 2017 | EEG and EEG reports | 3000 | NA | Diagnostic classification | A | NA | RNN and word2vec | F1 0.9277 |
| [[77]](https://paperpile.com/c/26vM9E/wghgn) | ICDM | III | 2016 | Proprietary EHR | 218, 680 | NA | concept embedding | C | 2011-2015 | CNN |  |
| [[66]](https://paperpile.com/c/26vM9E/0Kpd) | arxiv | III | 2017 | MIMIC III | 35,554 | NA | concept embedding | C | 2000-2011 | autoencoder | MSE ~0.005 |
| [[75]](https://paperpile.com/c/26vM9E/c8T5U) | arxiv | III | 2017 | EHR with clinical notes in MIMIC III | 25,000  visits | NA | Predict medication | B | 2000-2011 | CNN | Micro F1 0.65 |
| [[33]](https://paperpile.com/c/26vM9E/tsoL5) | BMC Medical Informatics and Decision Making | II | 2017 | Clinical notes from Rheumatology Clinic | 662 | NA | concept embedding for lupus patient | C | NA | word2vec | AUC 0.963 |
| [[76]](https://paperpile.com/c/26vM9E/6i10U) | JBHI | II | 2017 | Australian data | 4993 | NA | Readmission prediction | B | NA | CNN | ACC ~0.75 |
| [[28]](https://paperpile.com/c/26vM9E/SJ8aw) | KDD | III | 2017 | MIMIC III & Sutter Health | NA | 50,206 visits (MIMIC III) &  2,415,414 visits (Sutter Health) | Sequential prediction of medication | B | NA | RNN | Jaccard Coefficient ~ 0.55 |
| [[27]](https://paperpile.com/c/26vM9E/2vraf) | ICLR | III | 2016 | Vanderbilt EHR | 610,076 | NA | Sequential prediction of medication | B | NA | LSTM | micro-averaged  AUC 0.927 |
| [[83]](https://paperpile.com/c/26vM9E/0E6u2) | JBI | II | 2017 | Mount Sinai Health System | 377,686 | NA | Use vital sign and lab test to predict predict chronological age | A | NA | Autoencoders | na |
| [[99]](https://paperpile.com/c/26vM9E/uwAzr) | PSB | II | 2018 | EHR from Mount Sinai | 1,304,192 | NA | Concept embedding and cohort detection | C | NA | Word2vec | Auc 0.98 |
| [[88]](https://paperpile.com/c/26vM9E/dZP4c) | ACL | III | 2016 | Clinical text from Detect-HAI Corpus | 213 | NA | Concept embedding and association mining | C | NA | SAE and RBM | F-score 0.83 |
| [[53]](https://paperpile.com/c/26vM9E/xgIja) | MLHC | III | 2016 | CHB-MIT EEG data | 23 patients | 969 hrs recording | Seizure detection | A | NA | RCNN | Sensitivity 85% |
| [[89]](https://paperpile.com/c/26vM9E/TuoaY) | American Heart Associatio Circulation | I | 2017 | Geisinger Health System EHR | 200,539 | 414,006 | Concept embedding | C | NA | deep-autoencoder | 10.8% increase (from 0.52 to 0.58) in concordance score |
| [[84]](https://paperpile.com/c/26vM9E/sAVku) | IEEE TBME | II | 2017 | Acute coronary syndrome cohort EHR | 3,464 | NA | Disease classification | A | NA | regularized stacked denoising auto-encoder | AUC 0.868 |
| [[128]](https://paperpile.com/c/26vM9E/umDPr) | arXiv | III | 2017 | EHR data from California | 522,056 | NA | Suicide Classification | A | 2006-2009 | Fully connected DNNs | AUC 0.958 |
| [[22]](https://paperpile.com/c/26vM9E/qLcRf) | JBHI | II | 2017 | Breast and lung cancer pathology reports | NA | NA | Concept embedding | C | NA | CNN | micro-F score of 0.722 |
| [[60]](https://paperpile.com/c/26vM9E/YccHl) | MLHC | III | 2017 | Proprietary ICU data | NA | 51,697 admissions | Use physiological and medication data to predict onset of sepsis | B | NA | Gaussian process + RNN | Auc 0.85-0.9 |
| [[61]](https://paperpile.com/c/26vM9E/kCxMj) | ICML | III | 2017 | Proprietary ICU data | NA | 49,312 admissions | Use physiological and medication data to predict onset of sepsis | B | NA | Gaussian process + RNN | Auc ~0.98 |
| [[62]](https://paperpile.com/c/26vM9E/H7Y8z) | MLHC | III | 2017 | Data from PRAEGNANT study network | 4,357 | NA | Modeling survival time of breast cancer | B | 1961-2016 | RNN | Mean absolute error: 132.9, R-squared: 0.645 |
| [[68]](https://paperpile.com/c/26vM9E/WIiqS) | arXiv | III | 2017 | Proprietary clinical text | 115,232 | 2,735,6487 notes | Concept embedding of clinical notes | C | NA | RNNs |  |
| [[100]](https://paperpile.com/c/26vM9E/3feAL) | AAAI | III | 2017 | MIMIC III | 58,000 | NA | Concept embedding | C | NA | Memory networks | Auc ~0.767 |
| [[101]](https://paperpile.com/c/26vM9E/j4Mst) | arXiv | III | 2017 | Philips eICU | 17,693 | NA | Data generation | E | NA | GAN | NA |
| [[55]](https://paperpile.com/c/26vM9E/FC68x) | arXiv | III | 2017 | MIMIC III | NA | 30,931 sequences | Diagnosis based on imputation | A,D | NA | VAE+RNN | Mirco-AUC: 0.958 |
| [[69]](https://paperpile.com/c/26vM9E/PupBo) | BIBM | II | 2017 | EEG from UCI | NA | 600 | Concept embedding | C | NA | LSTM Autoencoders | multiple |
| [[78]](https://paperpile.com/c/26vM9E/D4Pl4) | NIPS ML4HC | III | 2017 | Private EHR | 218,680 | 14,969,489 | Concept embedding | C | NA | CNN | AUC 0.9 |
| [[93]](https://paperpile.com/c/26vM9E/iQg7I) | PSB | II | 2017 | PRO-ACT ALS | 1,824 | NA | Data imputation | D | 3-5 yrs | AE | NA |
| [[129]](https://paperpile.com/c/26vM9E/BGNTh) | Advances in Information Retrieval | III | 2016 | Mount Sinai EHR data | 105,000 | NA | Concept embedding | C,D | NA | DAE | multiple |
| [[92]](https://paperpile.com/c/26vM9E/k2JUA) | arXiv | III | 2017 | Breast cancer cohort | 569 | NA | Data generation | D | NA | Autoencoders+GANs | Accuracy 0.98 |
| [[91]](https://paperpile.com/c/26vM9E/37C0p) | ICDM | III | 2017 | CHB-MIT dataset, MIT-BIH arrhythmia database | 23, 48 | NA | Concept embedding | C | NA | SAE | AUC 0.9677(seizure detection); 0.9965 (arrhythmia detection) |
| [[63]](https://paperpile.com/c/26vM9E/h73Nk) | ICHI | II | 2017 | CIBMTR registry database | 6,021 | NA | predict expert action | B | NA | RNNs | Accuracy@5 ~0.941 |
| [[90]](https://paperpile.com/c/26vM9E/M49Aj) | arXiv | III | 2017 | University Hospital of North Norway | 883 | NA | Learn time series representation with missing data | C,D | 2004-2012 | AE | AUC 0.813 |
| [[96]](https://paperpile.com/c/26vM9E/qP66O) | JBI | II | 2015 | Clinical notes from annotated corpus of Swedish health records | NA | 700M tokens | ADE classification | A | NA | Distributed embedding |  |
| [[79]](https://paperpile.com/c/26vM9E/1x88l) | JAMIA | II | 2018 | i2b2/VA data and MIMIC III | NA | NA | Concept embedding | C | NA | CNNs | micro-F 0.82 |
| [[54]](https://paperpile.com/c/26vM9E/25uHs) | JBI | II | 2017 | i2b2/VA data and MIMIC III | NA | NA | Classification of medical problem-treatment relations | A | NA | LSTM | micro-F 0.815 |
| [[64]](https://paperpile.com/c/26vM9E/xOtNq) | MLHC | III | 2016 | Stage 4 CKD cohort | 5,484 | 29,937 | Multitask prediction of disease | B | NA | LSTM | NA |
| [[72]](https://paperpile.com/c/26vM9E/iZ16Y) | arXiv | III | 2018 | Proprietary EHR | NA | 50,128 notes | Multilabel classification of doctor notes | A | NA | CNN+residual net | AUC ~0.95 |
| [[73]](https://paperpile.com/c/26vM9E/L4Ldd) | arXiv | III | 2016 | NA | 298K patient | NA | Multilabel prediction of disease | A | NA | CNN | multiple |
| [[80]](https://paperpile.com/c/26vM9E/u9J1b) | NIPS ML4HC | III | 2016 | MIMIC III | 31,244 | 812,158 notes | Concept embedding for mortality prediction | C | NA | CNN | AUC 0.853~0.963 |
| [[130]](https://paperpile.com/c/26vM9E/9UT2H) | BIBM | II | 2017 | Stanford EHR | 2m | NA | Mortality prediction | B | 1995-2014 | DNN | AUC 0.93 |
| [[65]](https://paperpile.com/c/26vM9E/eTU5G) | arXiv | III | 2018 | EHR from Multiple institutes | 114,003 | 216,221 hospitalizations | Multiple medical event prediction | B | NA | LSTM | AUC 0.75~0.95 |

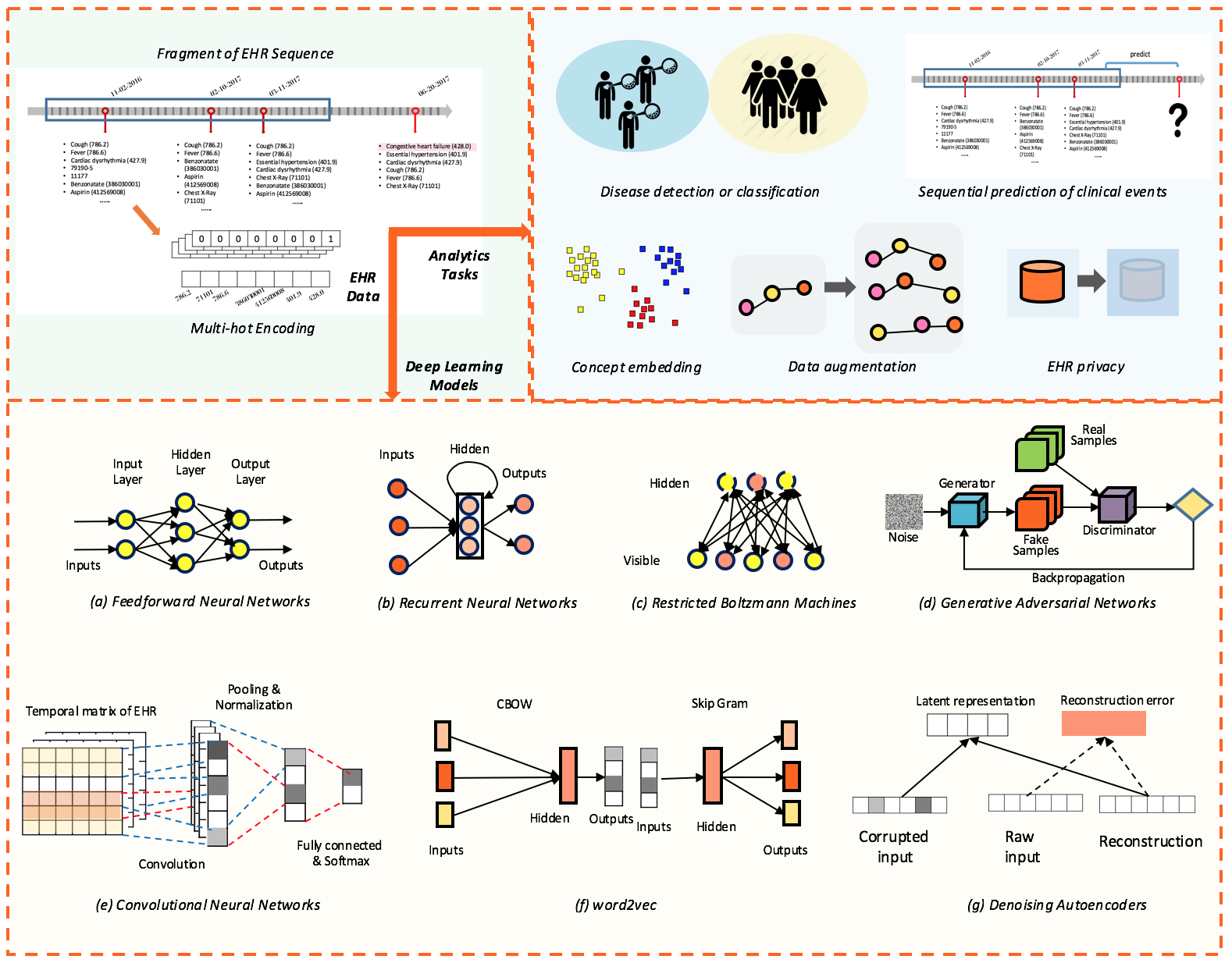
Table S2: Summary of challenges and solutions for articles selected in this literature review

|  |  |  |  |
| --- | --- | --- | --- |
| Author | Task | Targeting Challenge | Solutions |
| [[26]](https://paperpile.com/c/26vM9E/7NgMs) | Embedding & eval with prediction | How to predict the physician diagnosis and medication order of  the next visit, as well as the next visit time. | An end-to-end RNN based model to both learn efficient representation and make accurate multilabel prediction. |
| [[51]](https://paperpile.com/c/26vM9E/qFU1A) | Label clinical events (classification) from clinical notes | Extract exact clinical phrase is important since partial phrase could be a different disease; challenge of model long-term dependency. | Model CRF pairwise potentials using LSTM, and model approximate version of skip chain CRF to capture long-term dependency. |
| [[56]](https://paperpile.com/c/26vM9E/ADfYq) | readmission prediction | How to model temporal dependence and irregularity, as well as the effect of intervention | Built on LSTM, uses time parameterizations to handle time irregularity by moderating the forgetting and consolidation of memory of cells. Also model effects of intervention via gates. |
| [[45]](https://paperpile.com/c/26vM9E/dKEqV) | Multilabel classification to tag patient notes with ICD-9 codes | Codes are structured hierarchically; discharge notes can be very long and have large variance in note length. | Word vector representation with Glove, and then predict codes using LSTM. |
| [[14]](https://paperpile.com/c/26vM9E/OegXl) | Comp. phenotyping | Time irregularity | learns the similarity between two longitudinal patient record sequences through dynamically matching temporal patterns in  patient sequences |
| [[11]](https://paperpile.com/c/26vM9E/QPiJO) | Comp. phenotyping | Time irregularity | A subspace decomposition approach that learns a neural network that performs a decomposition of the cell memory into short and long-term memories. |
| [[64]](https://paperpile.com/c/26vM9E/xOtNq) | Multitask prediction of disease onset | Inductive predict disease for patients who did not have this disease | Multitask LSTM |
| [[115]](https://paperpile.com/c/26vM9E/KFV0n) | Causal inference | Treatment-effect analysis | Multitask deep neural networks |
| [[116]](https://paperpile.com/c/26vM9E/Mqph5) | Predictions of Cardiac Arrhythmia and Heart Failure | Clinical outcome prediction based on multiple data modalities | Use deep multitask learning approaches to jointly model different data modalities |
| [[117]](https://paperpile.com/c/26vM9E/bIZtv) | Multi-modal Comp. Phenotyping | Learn representation from multi-modal EHR data | Use Poisson Factor Analysis modules to model each modality, and then fused multimodalities via deep hierarchy of common hidden units. |
| [[118]](https://paperpile.com/c/26vM9E/3ooWq) | Clinical event prediction | Lack of label | Use transfer learning, trained dense and lower dimensional representation using autoencoders from source data, and then use these representations as features in a classifier trained to predict  targets task using a much smaller number of reliable labels. |
| [[20]](https://paperpile.com/c/26vM9E/dHdmC) | Multi-label clinical notes classification | Model transparency: need to highlight elements in documents that explain the labels in multi-label classification | Hierarchical attention bidirectional GRU approach to tag clinical notes with diagnosis codes. |
| [[43]](https://paperpile.com/c/26vM9E/uHurb) | Heart failure prediction | Model interpretability | Use attention model to generate attention weights for each clinical event in the past, providing interpretation as to which clinical events the model deems more important or less important for predicting disease onsets of future events. |
| [[44]](https://paperpile.com/c/26vM9E/os23r) | sequential & HF onset prediction | Model interpretability; knowledge incorporation | Learn the latent embedding of a clinical event as a convex combination of the basic embeddings of the event itself and its ancestors. The combination coefficients, namely the attention weights, were learned jointly with the RNN model parameters. |
| [[47]](https://paperpile.com/c/26vM9E/BdDti) | prediction | interpretability | It employs bidirectional recurrent neural networks to remember all the information of both the past visits and the future visits, and it introduces three attention mechanisms to measure the relationships of different visits for the prediction. |
| [[94]](https://paperpile.com/c/26vM9E/uuwSi) | Mimic learning for interpretation; ICU outcome prediction | Model interpretability | Mimic learning: A vector of class probabilities based on the results of a deep neural network was firstly assigned to each sample, and then used as target label to train an additional gradient boosting tree. |
| [[122]](https://paperpile.com/c/26vM9E/eYQCJ) | Mimic learning for interpretation | Model interpretability | Mimic learning: A vector of class probabilities based on the results of a deep neural network was firstly assigned to each sample, and then used as target label to train an additional gradient boosting tree. |
| [[12]](https://paperpile.com/c/26vM9E/m1Zpw) | CHF and COPD prediction | Temporal dependency of clinical events. | Layers with convolution filters on local features to capture higher-order event relations, additional convolutional operation over the temporal dimension to capture temporal patterns. |
| [[49]](https://paperpile.com/c/26vM9E/VWT57) | readmission prediction | Latent process from EHR is complex, challenges in capturing dynamic structure of care. | Take an algebraic view to embed discrete events into a continuous vector and then use regularized LSTM for prediction |
| [[30]](https://paperpile.com/c/26vM9E/Djp8) | Unsupervised embedding & prediction | Traditional supervised representation learning in the form of 2-d vectors are often sparse, noisy, and repetitive, | Used stacks of denoising autoencoders to process EHRs in an unsupervised manner that captured stable structures and regular patterns in the data. |
| [[31]](https://paperpile.com/c/26vM9E/UOdXC) | Embedding & eval with prediction | How to learn efficient representations of discrete high dimensional concepts with considering temporality, interpretability, and scalability. | Generate low-dimensional embedding of clinical events using a two-layer neural networks. |
| [[36]](https://paperpile.com/c/26vM9E/7JkL3) | predicting drug-Induced  laboratory test | multi-modality, mixing categorical and continuous data with semi-structured and free text medical notes. | generate continuous time series that display effects of exposure changes with a simple GAN architecture. |
| [[23]](https://paperpile.com/c/26vM9E/vfOZT) | Predict HF onset | Complex EHR structure, temporality, irregularity | RNN-based model to solve these challenges. |
| [[98]](https://paperpile.com/c/26vM9E/0tpNC) | Comp. phenotyping | The original RBM can have negative coefficient, also ignores explicit structures in EMR | Modify RBM with nonnegativity and structural smoothness constraints and embed medical concepts in a low dimensional vector space in unsupervised setting. |
| [[58]](https://paperpile.com/c/26vM9E/qmPSg) | Sequential prediction | Combine static information with temporal ones | Use feedforward neural nets as a wrapper outside of GRU |
| [[85]](https://paperpile.com/c/26vM9E/KB34) | Comp. Phenotyping  for vital signs | Difficult to model underlying complex and nonlinear relationship among observed  measurements | Use prior knowledge to regularize parameters in the topmost layers, and a scalable procedure for training neural networks of different sizes but with partially shared architectures. |
| [[10]](https://paperpile.com/c/26vM9E/C1B2) | Comp. Phenotyping | Lots of unlabeled data | Leveraged denoising autoencoders (DA) to enhance the prediction performance of a random forest (RF) model. |
| [[86]](https://paperpile.com/c/26vM9E/YQEj) | Comp. Phenotyping | Noisy, sparse, and irregular sampled | autoencoders with longitudinal probability densities inferred using Gaussian process regression. |
| [[37]](https://paperpile.com/c/26vM9E/bUprs) | Data de-identification | Need a reliable automated de-identification system | Used ANN for token and character embeddings, then automatically jointly learns the parameters for the embeddings, the bidirectional LSTMs as well as the label sequence optimization, and can make use of token embeddings pre-trained on large unlabeled datasets. |
| [[50]](https://paperpile.com/c/26vM9E/dBSm9) | Sequential labeling of clinical notes (for ADEs) | Noisy, incomplete, abundant abbreviation, jargons and variations; ADEs are rare. | Applied RNN models and leverage their ability in memorizing contextual information. |
| [[46]](https://paperpile.com/c/26vM9E/Q0xco) | Multilabel diagnostic classification using multivariate ICU time series | Varying length of record, irregular sampled data, and long-term dependencies between different events | Applied LSTM to solve these mentioned challenges. |
| [[11]](https://paperpile.com/c/26vM9E/QPiJO) | Comp. phenotyping | Time irregularity | A subspace decomposition approach that learns a neural network that performs a decomposition of the cell memory into short and long-term memories. |
| [[52]](https://paperpile.com/c/26vM9E/3DRbw) | Predict whether users will achieve their weight objectives | Fuse multiple types of variables | Multi-model LSTM, process each modality separately and allowing for information flow between them by way of recurrent cross-connection, then perform a shared hyperparameter optimization to fuse multiple modalities. |
| [[95]](https://paperpile.com/c/26vM9E/ehzzA) | Diagnostic classification | na | na |
| [[81]](https://paperpile.com/c/26vM9E/IMM6) | Disease prediction | Firstly, the high-dimensional features and imbalanced class distributions often restrict model performance; Secondly, it is hard to disentangle the salient integrated features from heterogeneous information. | Investigate the performance of pretrained neural networks on health datasets.  visualize the contribution of risk factors to hidden units. Both of the deep belief nets and stacked auto-encoder models give some interesting patterns that contain latent information of features. |
| [[127]](https://paperpile.com/c/26vM9E/ChMt3) | causal  inference (Counterfactual Inference) | Answering EHR based counterfactual question | Formulate this as a domain adaptation problem and introduce a regularization by enforcing similarity between the distributions of representation learned from different interventions. |
| [[82]](https://paperpile.com/c/26vM9E/oMBC) | Seizure detection | Utilize large unlabeled data for multi-channel based seizure detection | adopt stacked sparse denoising  autoencoders to unsupervisedly learn multiple features  by considering both intra and inter correlation of EEG channels, and then concatenate these features to perform seizure detections. |
| [[34]](https://paperpile.com/c/26vM9E/Th7Os) | EHR generation | Data labeling, generation, and privacy. | Generate EHR data via adversarial training |
| [[71]](https://paperpile.com/c/26vM9E/j16Bl) | Diagnostic classification of normal/abnormal  heart sounds | Multiple domains (time and frequency) of features | Ensemble approach that combines a classifier trained with time-frequency features and a deep-learning (CNN) classifier |
| [[21]](https://paperpile.com/c/26vM9E/6KdB5) | Automated pathology report annotation | Missing punctuation, clinical findings interspersed with explanations, complex information about multiple specimens mentioned throughout the report. | Use multi-task learning paradigm to improve upon the DNN performance |
| [[32]](https://paperpile.com/c/26vM9E/RbbUw) | Comp. phenotyping | na | na |
| [[57]](https://paperpile.com/c/26vM9E/oa3Iw) | readmission prediction | long-term  Dependencies and temporal irregularity | Built on LSTM and uses time parameterizations to  handle irregular timing by moderating the forgetting and  consolidation of illness memory |
| [[59]](https://paperpile.com/c/26vM9E/Mpz3g) | multimodal  Prediction for both onset and weaning of interventions | Real time prediction of clinical intervention based on noisy, sparse, heterogeneous and imbalanced ICU data. | Two architectures: LSTM with feature-level occlusions; CNN with filter/activation visualization. |
| [[74]](https://paperpile.com/c/26vM9E/gQDOz) | interictal epileptic discharges detection based on time series | Manual marking epileptic discharges is time-consuming and subjective/limited | A deep learning based epileptic discharge detector for EEG-fMRI |
| [[13]](https://paperpile.com/c/26vM9E/IVSIZ) | Sepsis detection | Temporal dependency | Feedforward neural networks using long short-term memory |
| [[15]](https://paperpile.com/c/26vM9E/LwWP0) | Seizure detection | Neurologists employ direct visual inspection to identify epileptiform abnormalities, which is time consuming and subjective/limited. | A 13-layer deep convolutional neural network (CNN) algorithm is implemented to detect normal, preictal, and seizure classes |
| [[67]](https://paperpile.com/c/26vM9E/PHnGt) | Comp. phenotyping | Time irregularity | learns the similarity between two longitudinal patient record sequences through dynamically matching temporal patterns in  patient sequences |
| [[70]](https://paperpile.com/c/26vM9E/d6Nk2) | Data imputation | Missing values in the EHR | Directly model missingness as a features- sequence of missingness indicator and use RNN to predict. |
| [[48]](https://paperpile.com/c/26vM9E/3putj) | Diagnostic classification | Infer conditions from EEG reports are often subjective, and have moderate accuracy. | automatically extract word-level features using word2vec  and report-level features using RNN and infer underspecified information from EHRs with DNN. |
| [[77]](https://paperpile.com/c/26vM9E/wghgn) | Comp. phenotyping | Measuring patient similarity from EHR sees the challenges from heterogeneous, longitudinal, and sparse data. | Deep model based on temporal matching |
| [[66]](https://paperpile.com/c/26vM9E/0Kpd) | Comp. phenotyping | Robust representations of human physiology are complicated, and contain many non-obvious dependencies between observed measurements. Moreover, challenges of modelling evolving clinical time series data: varying-length, irregularly sampled or has missing values | uses autoencoders for physiological time series signal reconstruction, use RNNs to model signals of varying-length. |
| [[75]](https://paperpile.com/c/26vM9E/c8T5U) | Predict discharge medication at the time of admission | Two challenges: 1) information limited to unstructured notes; 2) often two or more medications are prescribed. | A CNN model that takes an admission note as input and predicts one or more discharge medication. |
| [[33]](https://paperpile.com/c/26vM9E/tsoL5) | Comp. phenotyping for lupus patient | Phenotyping lupus from clinical notes | Apply word2vec bayesian inversion method |
| [[76]](https://paperpile.com/c/26vM9E/6i10U) | Comp. phenotyping | Detect regular clinical motifs from irregular episodic records | Use convolutional neural net to construct  local clinical motifs to stratify the risk |
| [[28]](https://paperpile.com/c/26vM9E/SJ8aw) | Sequential prediction of medication | Label dependency; label mapping | Use recurrent decoder to model label dependency; content-based attention for label instance mapping; reinforcement learning to fine-tune parameters. |
| [[27]](https://paperpile.com/c/26vM9E/2vraf) | multilabel prediction of medication | na | na |
| [[83]](https://paperpile.com/c/26vM9E/0E6u2) | Use vital sign and lab test to predict predict chronological age | Identify cohort | EMR-wide association analyses and GWAS were performed to characterize discordant patients. |
| [[99]](https://paperpile.com/c/26vM9E/uwAzr) | Concept embedding and cohort detection | There is no systematic evaluation against gold standard | Systematically evaluate against gold PheKB algorithm. |
| [[88]](https://paperpile.com/c/26vM9E/dZP4c) | Concept embedding and association mining | Noisy input: concept may be implicitly mentioned, may have abbreviation, etc | SAE and RBM for better emebdding. |
| [[53]](https://paperpile.com/c/26vM9E/xgIja) | Seizure detection | Pattern are extremely variable inter- and intra- patients | An RCNN model to learn spatially invariant patterns of seizure. |
| [[89]](https://paperpile.com/c/26vM9E/TuoaY) | Concept embeddin | High dimensionality | DAE |
| [[84]](https://paperpile.com/c/26vM9E/sAVku) | Disease classification | Complex patterns | two constraints are added on SDAE to make the reconstructed feature representations contain more risk information of patients |
| [[128]](https://paperpile.com/c/26vM9E/umDPr) | Suicide Classification based on EHR | Novel problem | Large scale experiment, promising AUCs. |
| [[22]](https://paperpile.com/c/26vM9E/qLcRf) | Concept embedding | na | na |
| [[60]](https://paperpile.com/c/26vM9E/YccHl) | Use physiological and medication data to predict onset of sepsis | 1, the exact time of sepsis onset is generally unknown, usually observed indirectly from lab tests, 2) patient heterogeneity; 3) need to leave ample time to inform onset | Transfer raw physiological data through a multitask Gaussian process, and then use RNN to predict. |
| [[61]](https://paperpile.com/c/26vM9E/kCxMj) | Use physiological and medication data to predict onset of sepsis | 1, the exact time of sepsis onset is generally unknown, usually observed indirectly from lab tests, 2) patient heterogeneity; 3) need to leave ample time to inform onset | Transfer raw physiological data through a multitask Gaussian process, and then use RNN to predict. |
| [[62]](https://paperpile.com/c/26vM9E/H7Y8z) | Predict physician’s decision | Multiple decisions are correlated | Encode patient history with RNN, then solve target correlation problem with tensor factorization. |
| [[68]](https://paperpile.com/c/26vM9E/WIiqS) | Concept embedding of clinical notes | High dimensionality, sparsity, complex structure, need easy-to-transfer representations | RNNs |
| [[100]](https://paperpile.com/c/26vM9E/3feAL) | Concept embedding | Free-text medical concept embedding, need comprehension of free text | Used memory network, a model from knowledge base, to preserve the hierarchy of features in the memory. |
| [[101]](https://paperpile.com/c/26vM9E/j4Mst) | Generate realistic and labelled medical time series | Generate multivariate medical time series for various ICU settings | GAN with differential privacy |
| [[55]](https://paperpile.com/c/26vM9E/FC68x) | Imputation based diagnosis | Missing value; complex multivariate time series structure | Combine generative learning with discriminative learning |
| [[69]](https://paperpile.com/c/26vM9E/PupBo) | Concept embedding | With missing observations | LSTM autoencoder |
| [[78]](https://paperpile.com/c/26vM9E/D4Pl4) | Concept embedding | Local temporal dependency | CNN embedding |
| [[93]](https://paperpile.com/c/26vM9E/iQg7I) | Data imputation | Some missing at random; some not missing at random | Impute using autoencoders |
| [[129]](https://paperpile.com/c/26vM9E/BGNTh) | Concept embedding | NA | DAE |
| [[92]](https://paperpile.com/c/26vM9E/k2JUA) | Data generation | NA | NA |
| [[91]](https://paperpile.com/c/26vM9E/37C0p) | Concept embedding of continuous time series data | Lack of method to embed both discrete sequence and continuous time series | learns both inherent and embedding representations of biosignals |
| [[63]](https://paperpile.com/c/26vM9E/h73Nk) | predict expert action | NA | NA |
| [[90]](https://paperpile.com/c/26vM9E/M49Aj) | Representation learning for time series | Missing data | Use autoencoders |
| [[96]](https://paperpile.com/c/26vM9E/qP66O) | ADE classification based on distributed embedding | NA | NA |
| [[79]](https://paperpile.com/c/26vM9E/1x88l) | Concept embedding | NA | NA |
| [[54]](https://paperpile.com/c/26vM9E/25uHs) | Classification of medical problem-treatment relations | Temporal relation | LSTM |
| [[64]](https://paperpile.com/c/26vM9E/xOtNq) | Multitask prediction of disease onset | Inductive predict disease for patients who did not have this disease | Multitask LSTM |
| [[72]](https://paperpile.com/c/26vM9E/iZ16Y) | Multilabel classification of doctor notes | Diagnoses are correlated | Employ deep residual net on top of CNN encoder to capture diagnosis dependencies and incorporate information directly from the encoded sentence vector. |
| [[73]](https://paperpile.com/c/26vM9E/L4Ldd) | Multilabel prediction of disease | Data are multiresolution, observation are sparse | Impute with multivariate kernel regression;  take both imputed signal and binary observation as input, learn patterns with CNN. |
| [[80]](https://paperpile.com/c/26vM9E/u9J1b) | Concept embedding for ICU mortality prediction | complex multi-word or multi-sentence patterns | Use CNN to capture complex patterns |

**Figures**



**Fig. 1 Illustration of literature search and selection procedure.**



**Figure 2: Transform longitudinal EHR data into input vectors using multi-hot encoding (top left), which could support 5 categories of analytics tasks described in the survey (top right). The enablers of these tasks are different types of deep learning models (bottom). Bottom (a): Feedforward neural networks use multiple layers of fully-connected neural networks and non-linear activations (e.g. sigmoid or rectified linear unit). (b): Recurrent neural networks can process variable-length input sequence using its recurrent connection. (c): Restricted Boltzmann Machines are bipartite neural networks that consist of binary stochastic nodes. They can capture the latent representation of the input data by learning their generative probability. (d): Generative adversarial networks can generate realistic synthetic samples by putting the generator and the discriminator in an adversarial game. (e): Convolutional neural networks capture local features of the input data, and build those features up via a sequence of convolution to derive global features. (f): Word2vec exploits the co-occurrence information of discrete concepts (e.g. words in text, codes in EHR data) to derive their distributed representations. (g): Denoising autoencoders try to reconstruct raw input from its corrupted version, thus learning robust representations of the input data.**