Deep learning prediction models based on EHR trajectories: A systematic review

Ali Amirahmadi[a], Mattias Ohlsson[a,b], Kobra Etiminani[a]

[a]Center for Applied Intelligent Systems Research, Halmstad University

[b]Computational Biology & Biological Physics, Department of Astronomy and Theoretical Physics, Lund University

Background and motivations

Electronic health records (EHRs) are generated at an ever-increasing rate. EHR trajectories, the temporal aspect of health records, facilitate predicting patients' future health-related risks. It enables healthcare systems to increase the quality of care by early identification and primary prevention. Deep learning techniques have shown great capacity for analyzing complex data and have been successful for prediction tasks using complex EHR trajectories.

This systematic review aims to analyze recent studies to identify challenges, knowledge gaps, and ongoing research directions. Table1 shows an example of a patient´s EHR sequence data.

Patients demographic information:
Gender: M,
Date_of_birth: 1978-04-23,
Date_of_death:2018-02-10,
Religion: Catholic
Ethencity:White

Visit number: Visit 1 → 41 days → Visit 2 → 13 days → Visit 3 → 134 days → Visit 4 → 52 days → Visit 5 → 23 days → Visit 6 → 37 days → Visit 7

Visit number	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
Diagnosis	ABD. PAIN	GASTROPARESIS,HYPERTENSION	MALIGNANT HYPERTENSION	GASTROPARESIS	NAUSEA;VOMITTING; HYPOTENSION	HYPERTENSIVE URGENCY	UNCONTROLLED HYPERTENSION-NAUSEA-VOMITING
Medication	Clonidine TTS 1 Patch	Insulin, Cefazolin, D5W, Metoclopramide, Lorazepam, Clonidine TTS 1 Patch, Ondansetron, Heparin, Nitroprusside Sodium	Clonidine TTS 1 Patch, Heparin, Lorazepam, Metoclopramide, Magnesium Sulfate, Hydralazine HCl	Insulin, Heparin Flush Port, Acetaminophen, Multivitamins, NIFEdipine, Nitroglycerin Ointment 2%, Dolasetron Mesylate	Heparin, Hydralazine HCl, Ondansetron, Metoprolol, Promethazine HCl, Lorazepam	Heparin, Insulin Human Regular, Aspirin, Heparin Flush CVL (100 units/ml), Insulin, Dextrose 50%	Lisinopril, Sevelamer, Heparin Flush Port (10units/ml), Calcium Acetate, Pantoprazole, Atorvastatin
Procedures	Venous catheterization, Parent infus nutrit sub	Hemodialysis, Debridement of nail, Venous catheterization	Arterial catheterization	Hemodialysis, enous catheterization	Hemodialysis	Hemodialysis	Insert vasc access dev
Lab test	Urea Nitrogen:26_abnormal, Anisocytosis:2, Basophils:0.8	White Blood Cells:6.5, Anion Gap:13	Glucose:285_abnormal, Magnesium:1.5_abnormal	Hematocrit:34_abnormal	Hematocrit:39_abnormal, Hemoglobin:10_abnormal	White Blood Cells:6, Chloride:100	Glucose:172_abnormal, Creatinine:7.1_abnormal, Chloride:100
Clinical Text	Social History: Denies alcohol or tobacco use. Family History: His father recently died of ESRD and diabetes. you should take all your medications as directed. return immediately to the ER for any chest pain	Discharge Instructions: You were admitted to the hospital for abdominal pain and high blood pressure. You were treated with in intravenous pain medication which was converted to pills	Local anaesthesia was provided by lidocaine spray. No TEE related\complications. 0.2 mg of IV glycopyrrolate was given as an antisialogogue prior to TEE probe insertion	Sinus rhythm. Compared to the prior tracing of [**2187-1-19**] further evolution of the anteroseptal, lateral and apical myocardial infarction. Clinical correlation is suggested. TRACING #2	The central line tip is in the right brachiocephalic vein or proximal most SVC. The cardiac and mediastinal contours are stable. The lungs are clear. The lateral costophrenic angles are sharp without effusion.	38 year old man with recurrent episodes of abdominal pain and hypertension	increase in bilateral perihilar opacities is suggesting progression of pulmonary edema, although there is no sizable right or left pleural\n effusion
Admission and marital information	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private	Marital_status:Single, Admission_type: Emergency, Discharge_Location:Home, Insurance: Private

Table1: An example of a patient´s EHR sequence data and its different parts from MIMIC-III dataset.

Materials and Methods

- Searched the papers´title, abstract and keywords in September 2021
- Databases: Scopus, PubMed, IeeeExplore, the Association for Computing Machinery (ACM) Digital Library
- Phrases include in the query: patient health trajectory, EHR trajectory, patient care pathway, patient health pathway, patient trajectory prediction, disease prediction, deep learning, neural network, electronic health record, electronic health data, HER, electronic medical record and healthcare informatics
- We added 21 new relevant papers from references and authors searches.
- Fig1 shows the filters that applied on the resulting 119 papers
- We analyzed papers by investigating their primary challenges and solutions, suggested architecture, contributions, baseline models metrics, used data types, dataset and its availability.

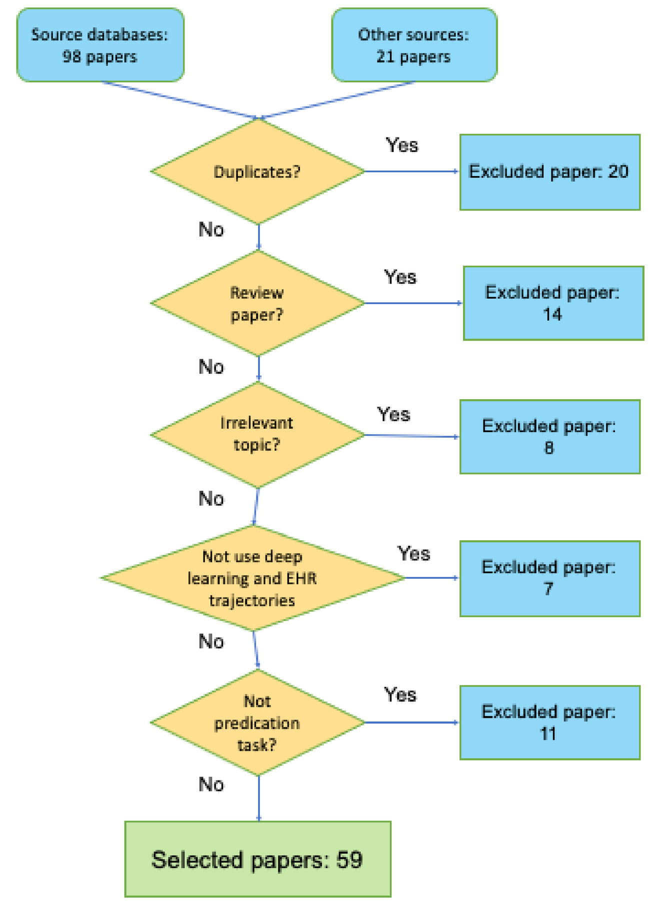

Fig1: Selected papers inclusion and exclusion flowchart.

Results

Publication characteristics

- After 2018, the number of publications has grown about three times from 6 to 17 (fig 2).
- The main reason for this growth could be the rise in EHR data availability rate and progress in deep learning methods and their availability to address data complexity.

Predicted subjects

In order to improve health care service quality, there are many areas where AI can be useful. We reviewed most frequent topics in table 2.

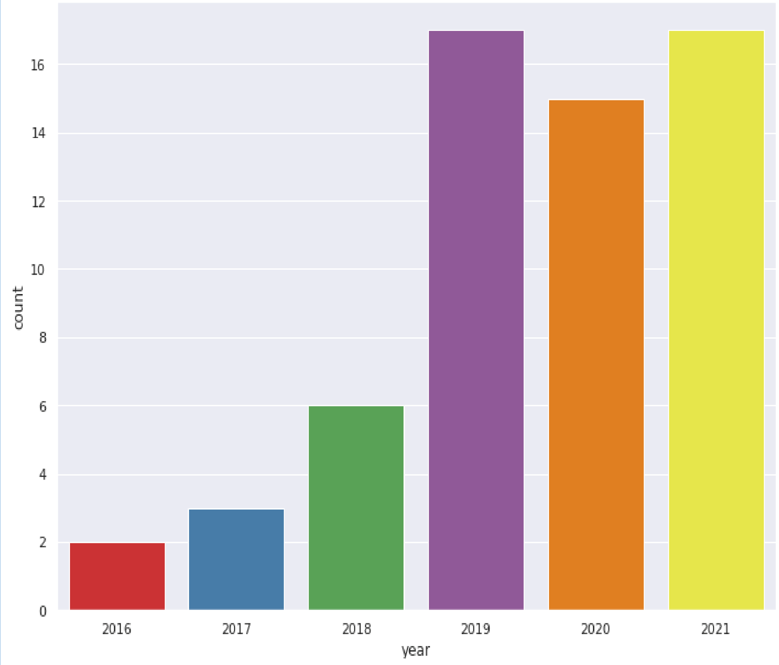

Fig2: Number of publications per year in collected papers.

Type of predicted outcome	Outcome	Number of papers
Disease prediction	All the diseases in the next visit	18
	Cardiovascular disease	18
	Diabetes	7
	Kidney diseases	7
	Chronic Obstructive Pulmonary Disease	4
	Alzheimer's diseases	3
Health-related prediction	Mortality prediction	5
	Readmission prediction	5

Table2: The most frequent predicted topics in reviewed papers.

Challenges and solutions

Predicting patients' future health-related risks is a complex task that is full of obstacles and difficulties. We review the most important challenges and suggested solutions as follow:

- **Heterogeneous data**
- ✓ EHRs consist of heterogeneous data sources with multiple data modalities
- ✓ Each one intends to record specific aspects of patients' health status and different methods have used to extract information from these modalities. Fig3 Shows an summary of extracting information from different modalities workflow and some of corresponding references.

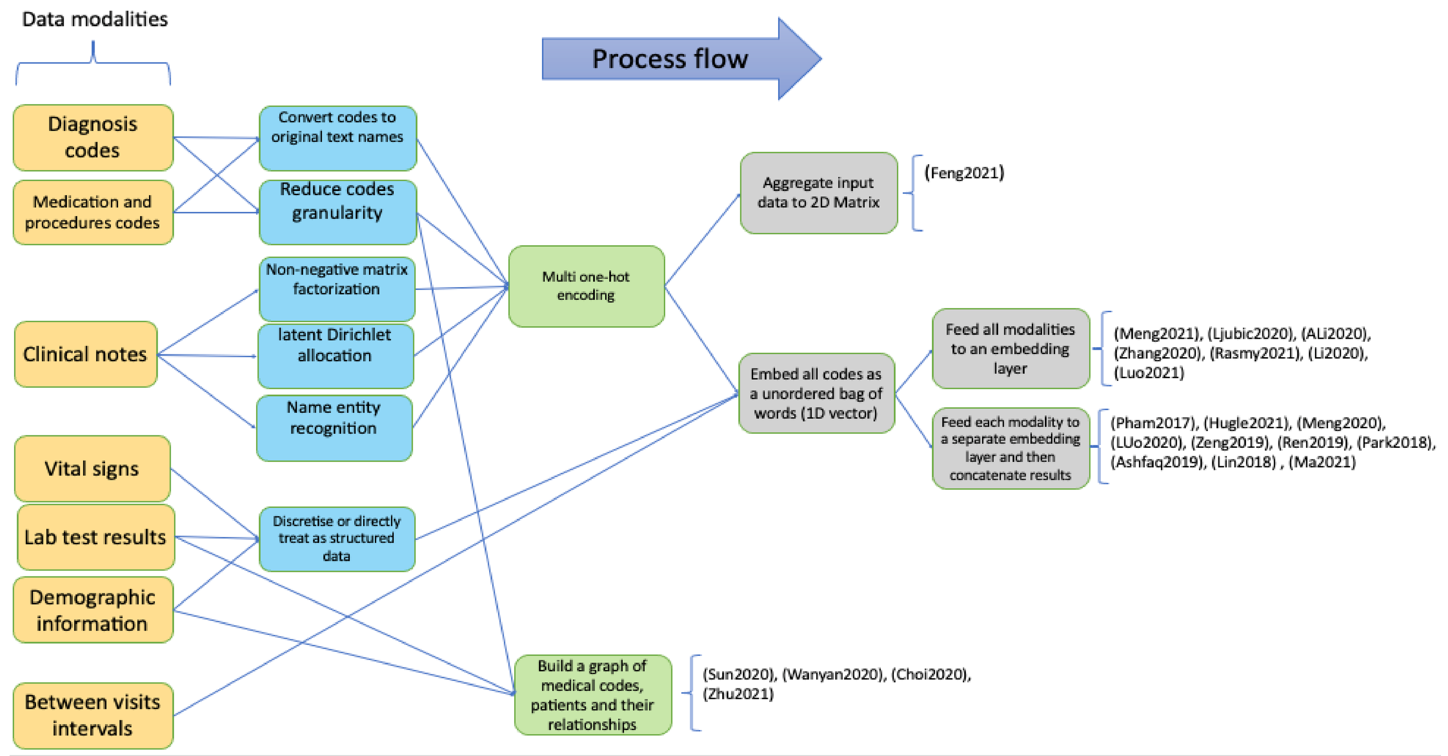

Fig3: Used methods to handle data heterogeneity in revied papers.

- **Representation of admission information**
- ✓ EHRs often have high dimensionality, sparsity and contain a large number of inconspicuous relations.
- ✓ In the dynamic approach, networks are trained end-to-end and usually in a supervised manner, while in the static approach, we train a separate network for data representation, e.g. using unsupervised learning
- ✓ Co-occurrence based methods build representation by considering relations within a single visit while sequential occurrence focuses on relations in neighbour visits (Fig 4)

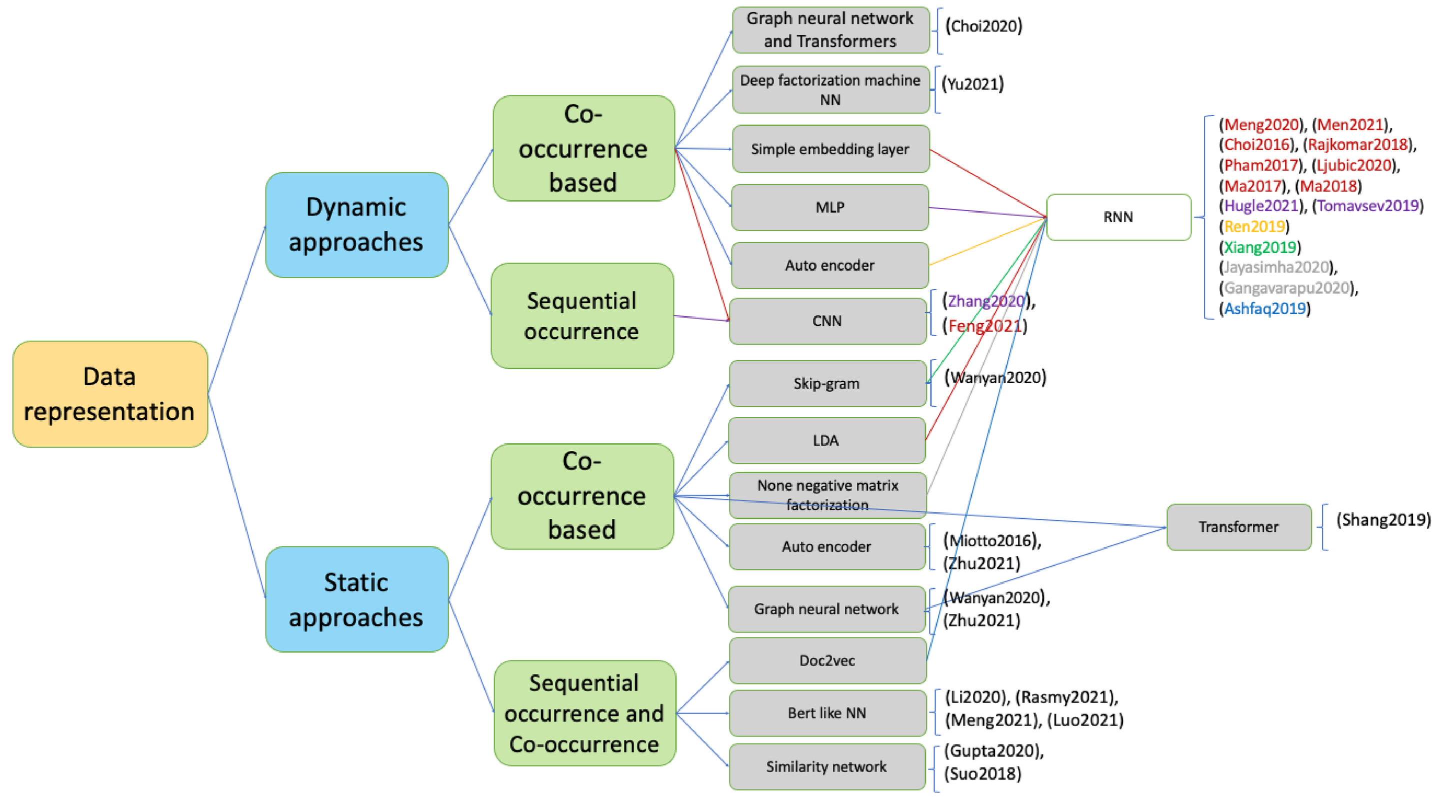

Fig4: Different data representation methods in reviewed papers.

- **Long term dependencies and irregularity in time**
- ✓ EHR data are longitudinal by nature where current state of a patient typically depends on data for previous visits.
- ✓ Considering the information of past visits is essential.
- ✓ To this end, using RNN, CNN and attention mechanism were the most common ways (Fig5).

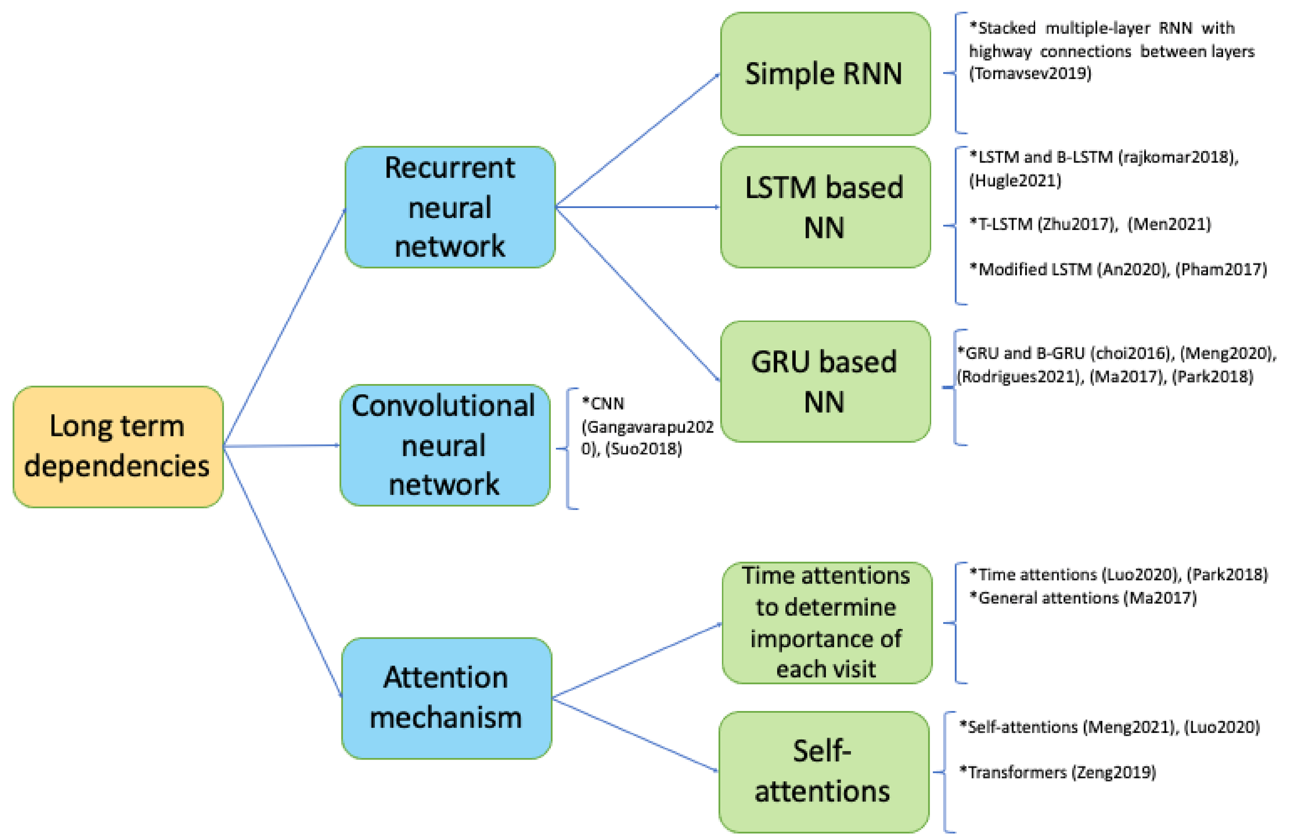

Fig5: Different ways to deal with long term dependencies in reviewed papers.

- **Interpretability and explainability**
- ✓ Explainability of models' decisions will be essential for acceptance among intended users.
- ✓ Here we found methods such as gradient based localization, visualizing attention weights and visualizing embedded feature space, in reviewed papers (Fig8).

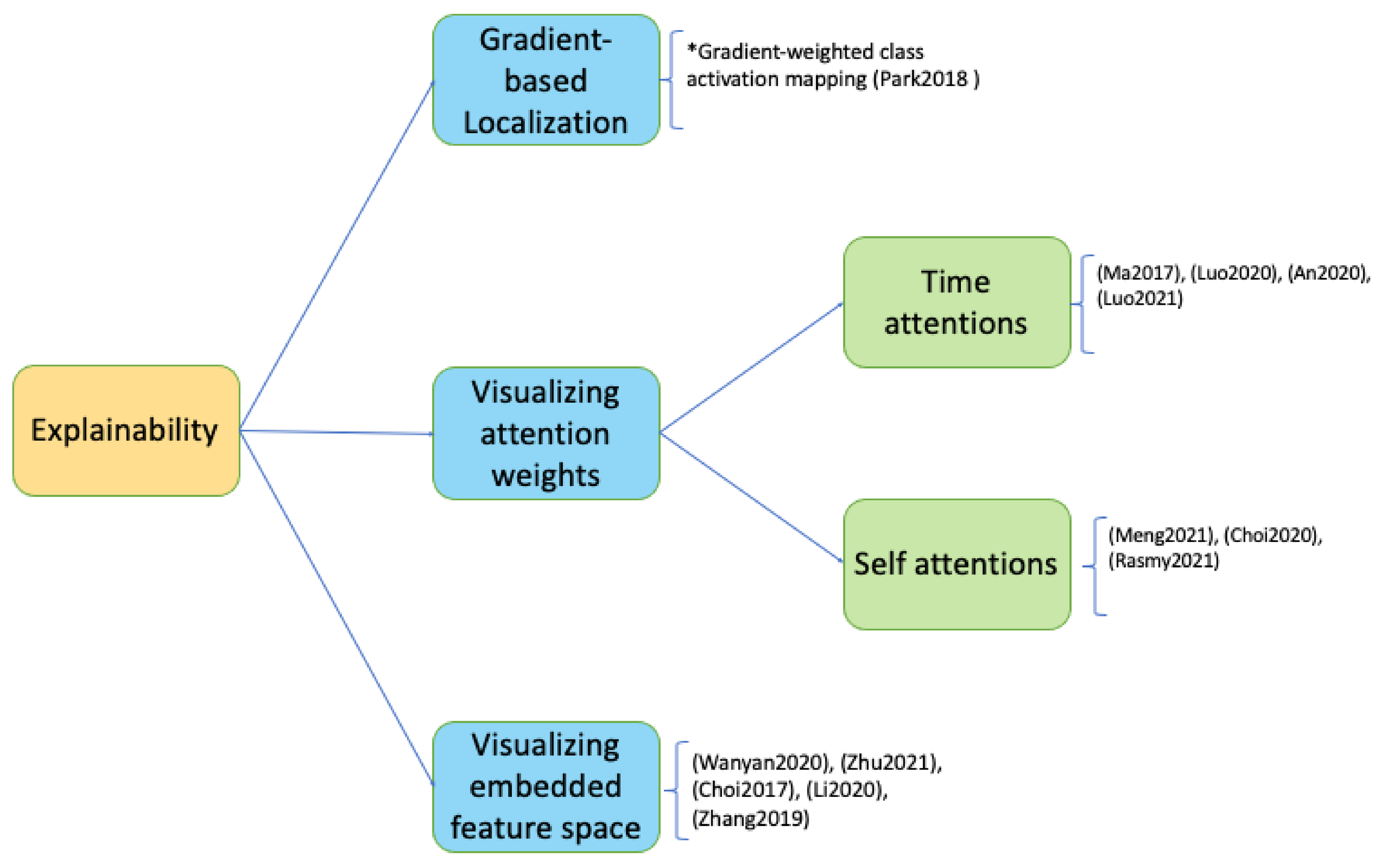

Fig6: Explainability methods that used in reviewed papers.

Conclusion

Identified challenges and directions of future research:

- ✓ Fully handling all aspects of EHR data in a single model is still challenging.
- ✓ There is a need to further develop models that can use both structured and unstructured data of different sources effectively.
- ✓ Representation methods should consider inter and intra dependencies between and within visits to make more accurate models.
- ✓ Handling the effect of rare diseases is another obstacle to model EHR data.
- ✓ Benchmark datasets to evaluate the accuracy and degree of explainability of NN models in this field is required.
- ✓ The privacy and fairness of EHRs predictive models need more investigations.