UniMed Multimodal Multitask Learning for Medical Predictions

1. 논문 소개

2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)

UniMed: Multimodal Multitask Learning for Medical Predictions

Xiongjun Zhao[1], Xiang Wang[2] Fenglei Yu[2] Jiandong Shang[3] Shaoliang Peng[12*]
[1]College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
[2]The Second Xiangya Hospital, Central South University, Changsha, China
[3]HeNan Supercomputer Center, ZhengZhou University, ZhengZhou, China
{xiongjunzhao, slpeng}@hnu.edu.cn, wangxiang@csu.edu.cn, shangjiandong@zzu.edu.cn

***Abstract*—Recently, deep learning techniques based on electronic health record (EHR) data have achieved success in medical prediction. However, due to the complexity, heterogeneity nature of EHR data, most previous studies build models based on single-modal data (e.g. the structured data or the unstructured free-text data). Although some studies have trained the models based on multimodal EHR data and achieved more advanced performance, they still suffer from the clinical practicability problems, as they require separate modeling for each medical prediction task. Moreover, they ignore the potential correlation between clinical prediction tasks. In this work, we propose UniMed, a *Uni*fied model handles multiple *Med*ical prediction tasks simultaneously by learning from multimodal EHR data. Our UniMed model encodes each input modality separately and uses a transformer decoder followed by task-specific prediction heads to predict each medical task. Experimental results conducted on publicly available EHR dataset demonstrate that there is a time-progressive correlation between medical prediction tasks and show the effectiveness of our method.**

***Index Terms*—Medical Predictions, Electronic Health Records, Multimodal, Multitask Learning**

I. Introduction

In recent years, deep learning techniques have achieved great performance in the medical domain, such as chest X-ray pneumonia diagnosis [1], skin cancer classification [2]. At the same time, with a large amount of patient-level EHR data being generated in hospitals every day, researchers began to focus on using EHR data to predict patient clinical outcomes, such as intensive care unit (ICU) mortality prediction [3], length-of-stay [4], and acute respiratory failure (ARF) prediction [5].

Unlike medical images and medical plain text data, real-world EHR data is longitudinal and multimodal, including structured data and unstructured data. For instance, Structured data in EHR can usually be divided into two modalities: time-invariant and time-dependent. Time-invariant data refers to static or discrete categories during hospitalization, such as patient gender, age and medication codes. Time-dependent data refers to dynamic or continuous features, such as vital signs and laboratory tests. Unstructured data refers to free-text, such as clinical notes. Therefore, the EHR data of patients can be summarized into the above three modalities.

Due to the complexity of EHR data, researchers only focused on structured data or unstructured free text data in previous work. Both [6] and [7] use time-dependent data to train a recurrent neural networks (RNN) for medical prediction. While [8] use transformer to model clinical notes for predicting hospital readmission.In general, most literature uses single-modality data as input to predict a single task. Recently, some research works have begun to enhance model prediction performance by fusing multimodal EHR data [9], [10]. Similarly, [11] propose multimodal fusion architecture search strategy for better leverage multimodal EHR data.

Despite the above achievements in medical prediction using deep learning for specific tasks, there has not been much effort to explore the potential correlation between medical tasks to improve prediction performance with multimodal EHR data. But in a clinical research, [12] shows that acute respiratory failure is related to a mortality rate of 35%-46% in clinic. Additionally, applying previous models to real clinical practice is a challenge, because it is necessary to build and train a model for each medical prediction task, even if their training data come from the same EHR system. Overall, as a step towards general intelligence, is it possible to use multimodal EHR data to build a unified single model that handles multiple tasks simultaneously, and take advantage of the potential correlation of medical tasks.

Inspired by the fact that physicians consider multiple data and previous clinical events when making decisions, we define the correlation between medical prediction tasks as a time-progressive, which means that the probability of the current task output can refer to the output of all previous tasks. As a step forward, we use multimodal multitask learning and build a simple but effective medical prediction model, UniMed. It takes three modalities of structured data and unstructured data as inputs and jointly train on multiple tasks. UniMed consists of three encoders that encode each input modality as a feature vector, and a transformer decoder over the encoded input modalities, and then applies task-specific prediction heads to the hidden states of the decoder to make the final prediction for each medical task. The self-attention mechanism of transformer can learn to focus on a main modality of the input for different tasks and consider previous outputs in current step. Compared to previous work on single-task learning, we train UniMed and achieve comparable performance to well-established prior work on variety of medical prediction tasks. Further analysis of the publicly available EHR dataset

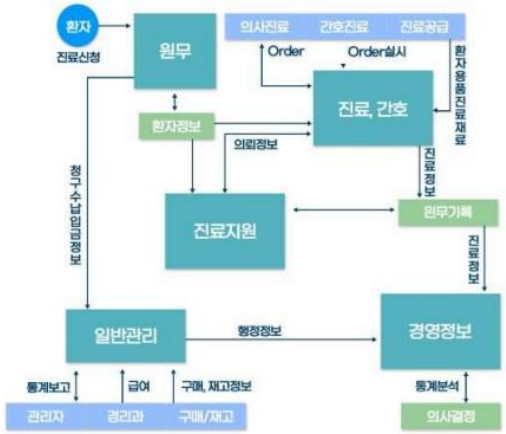
* Corresponding author

- **UniMed Multimodal Multitask Learning for Medical Predictions**
  - **발행년도**: 2022
  - **저자**: Xiongjun Zhao
  - **저널**: IEEE
  - **인용수**: 31
  - **요약**

EHR 데이터의 복잡성과 이질성으로 대부분의 이전 연구는 단일 모달 데이터를 기반으로 하며, 각 의학 예측 작업에서 별도의 모델링을 필요로함. 해당 논문은 멀티 모달 EHR 데이터를 활용하여 여러가지 예측을 한 번에 시도

2. Electronic Health Record

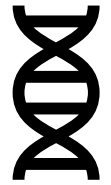
- EMR(Electronic Medical Record)
  - 진료기록을 수기로 관리하던 기존 시스템에서 벗어나, 컴퓨터에 입력하고 자료는 데이터 베이스로 관리
  - 의사중심으로 의학 자료의 수집과 관리 그리고 이용이 이루어지며, 주로 의료기관 내에서만 활용
- EHR(Electronic Health Record)
  - 의료정보의 범주가 병원에서 개인 건강에 관한 모든 정보로 확대됨
  - 의료기간 간의 정보공유가 이루어지는 등 의료정보가 병원 내에서만 유통되지 않고 병원 외부에서도 활용됨
  - 데이터의 형태와 크기가 기하급수적으로 증가
  - 공유를 고려하지 않고 구축된 각 병원간 데이터의 형태가 상이하여 표준화의 필요성이 제기됨



2. Electronic Health Record

- ADMISSIONS.csv.gz
- CALLOUT.csv.gz
- CAREGIVERS.csv.gz
- CHARTEVENTS.csv.gz
- CPTEVENTS.csv.gz
- DATETIMEEVENTS.csv.gz
- DIAGNOSES_ICD.csv.gz
- DRGCODES.csv.gz
- D_CPT.csv.gz
- D_ICD_DIAGNOSES.csv.gz
- D_ICD_PROCEDURES.csv.gz
- D_ITEMS.csv.gz
- D_LABITEMS.csv.gz
- ICUSTAYS.csv.gz
- INPUTEVENTS_CV.csv.gz
- INPUTEVENTS_MV.csv.gz
- LABEVENTS.csv.gz
- LICENSE.txt
- MICROBIOLOGYEVENTS.csv.gz
- NOTEEVENTS.csv.gz
- OUTPUTEVENTS.csv.gz
- PATIENTS.csv.gz
- PRESCRIPTIONS.csv.gz
- PROCEDUREEVENTS_MV.csv.gz
- PROCEDURES_ICD.csv.gz
- README.md
- SERVICES.csv.gz
- SHA256SUMS.txt
- TRANSFERS.csv.gz

환자정보
전실 및 입/퇴원 기록(중환자실, 일반병동, 응급실)
처치 기록
투여 기록
수술 기록
간호 기록
진료 기록
영상 기록
랩 검사 기록



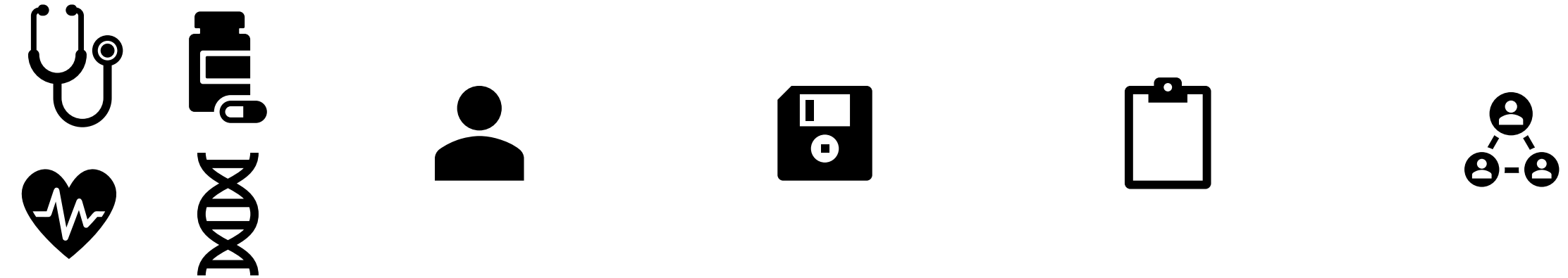
유전자 정보

라이프로그

소셜

2. Electronic Health Record

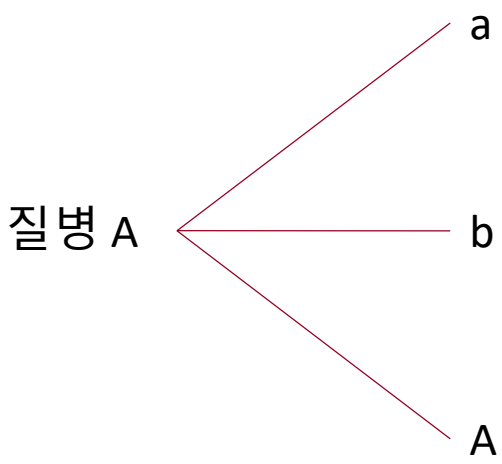
- Multi Modal

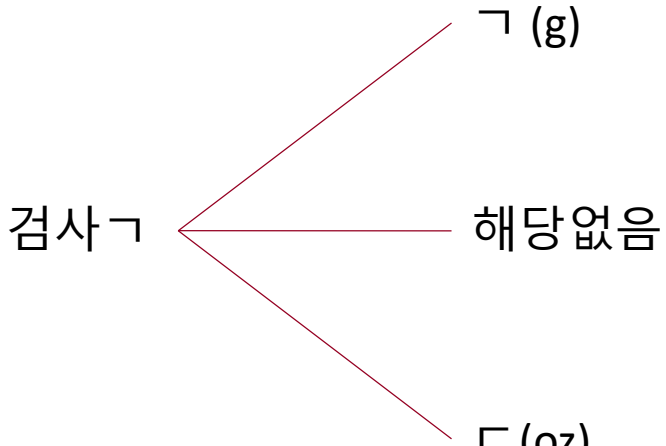


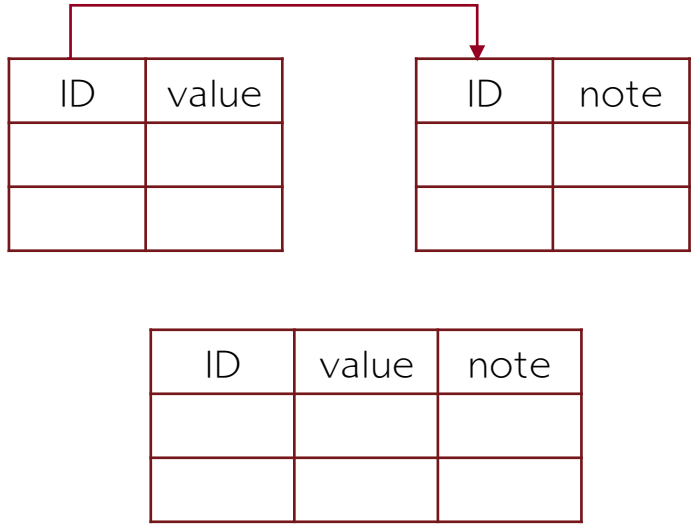
2. Electronic Health Record

◦ **병원간 이질성**

용어	검사/진단	데이터 저장

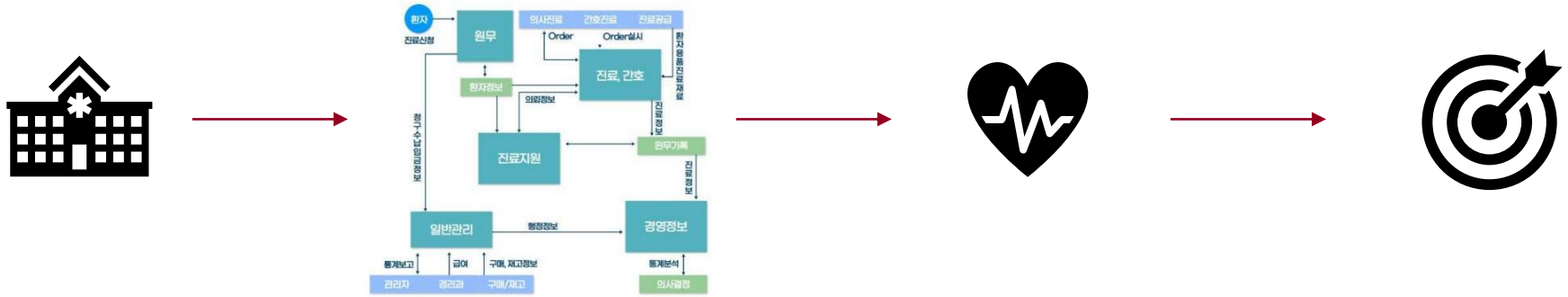






3. Objective

◦ 기존 연구들



병원 A

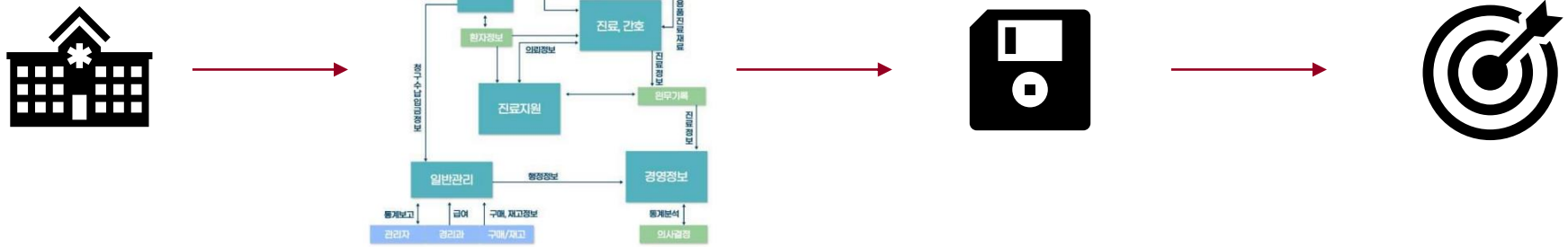
EHR

Modal

Object

3. Objective

◦ 기존 연구들



병원 A

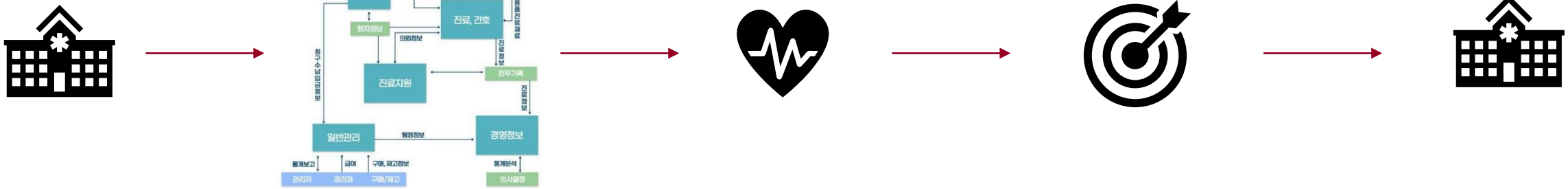
EHR

Modal

Object

3. Objective

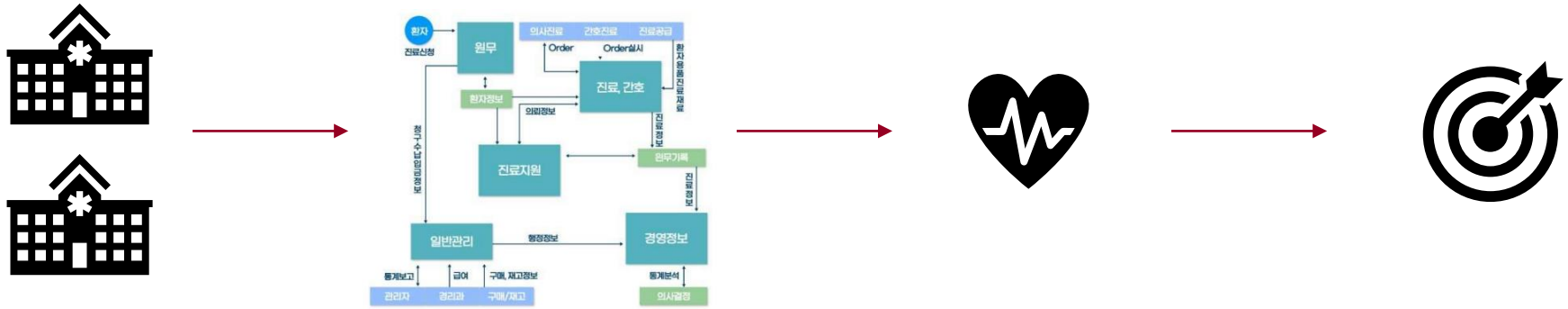
◦ 기존 연구들



병원 A | EHR | Modal | Object | 병원 B

3. Objective

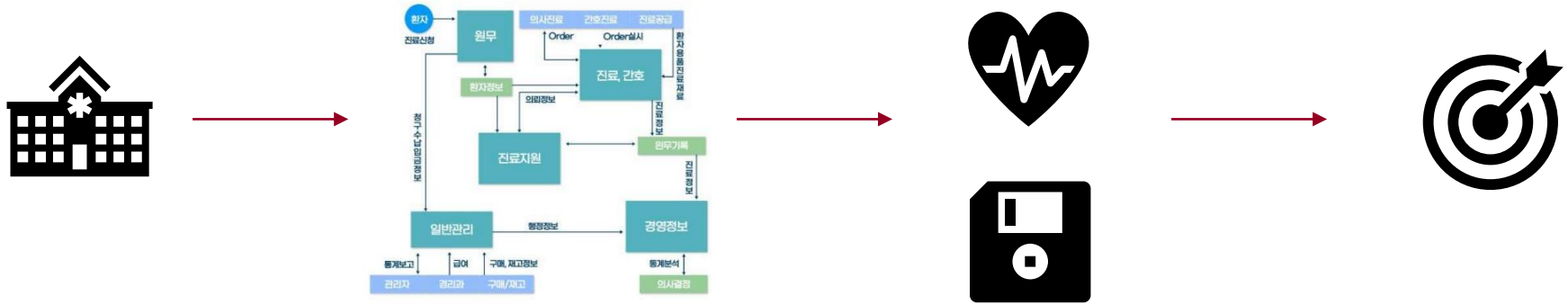
◦ 기존 연구들



병원 A, B → EHR → Modal → Object

3. Objective

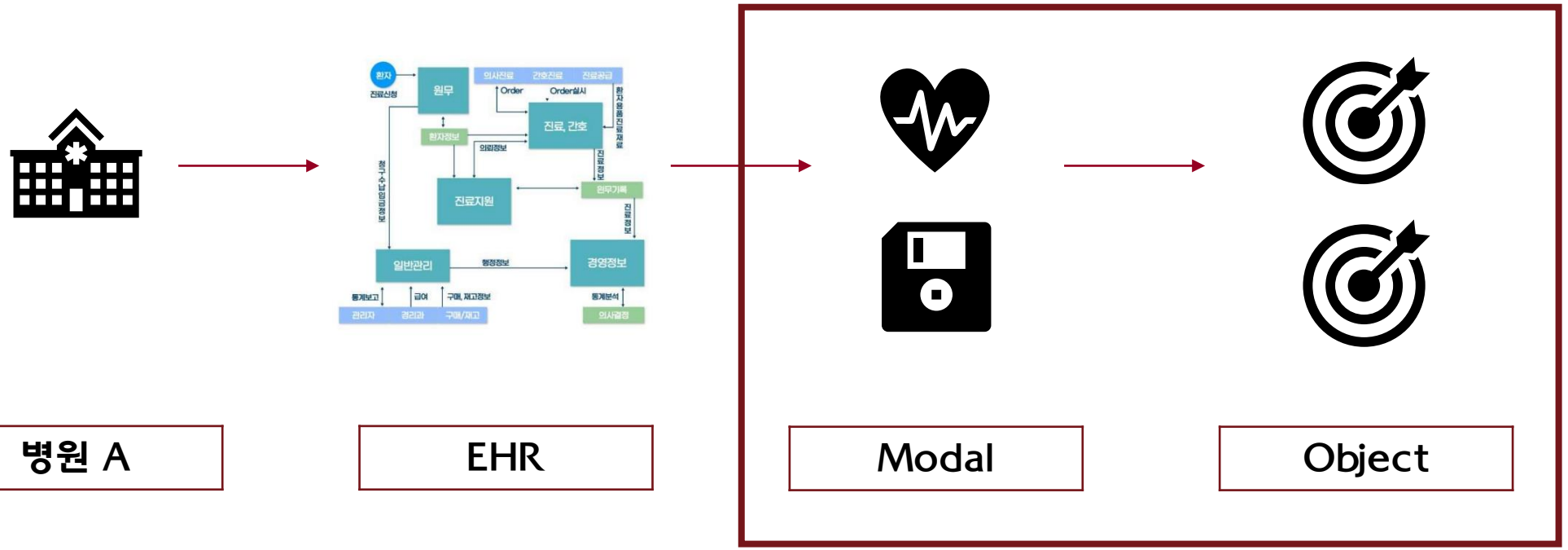
◦ 기존 연구들



병원 A → EHR → Modal → Object

3. Objective

◦ 논문의 목적



3. Objective

- **Inductive Bias**
  - 주어지지 않은 입력의 출력을 예측하는 것

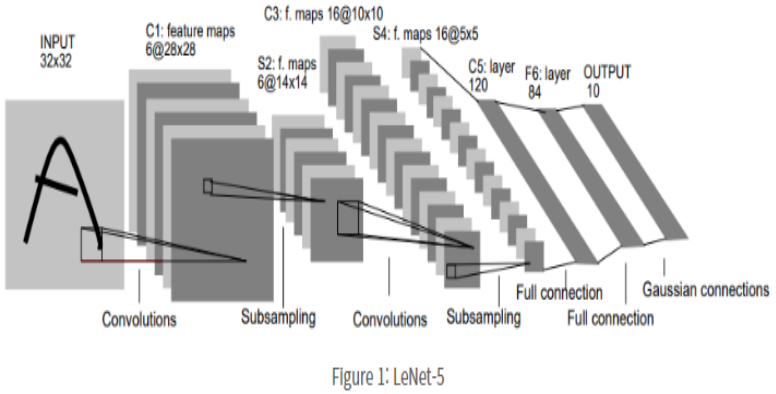
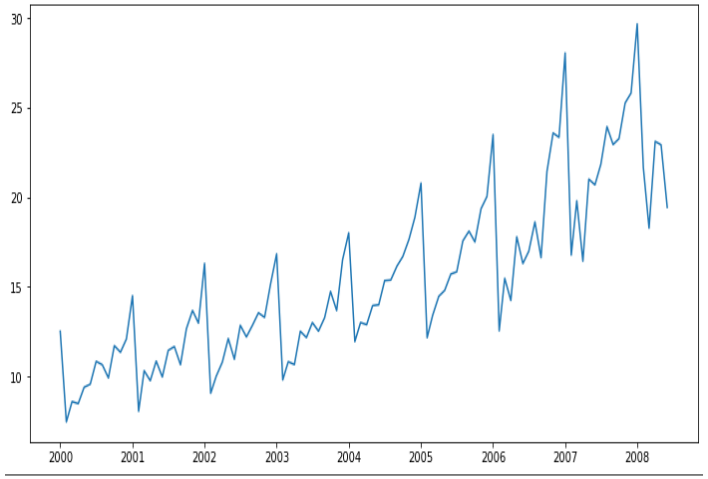
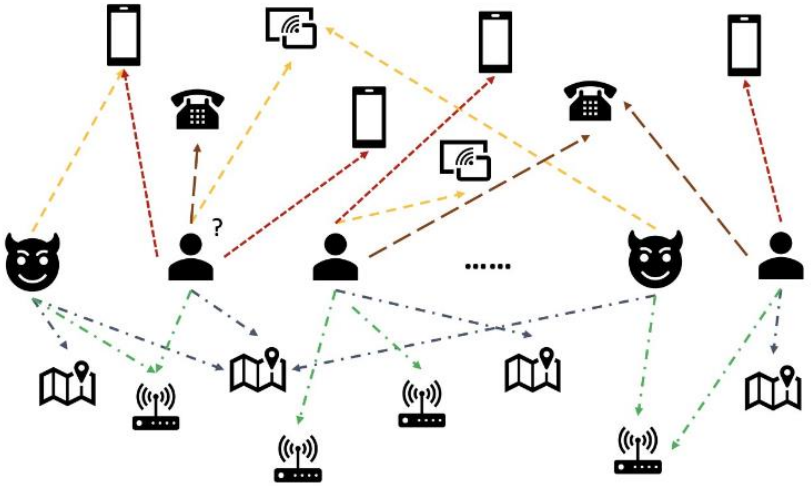


Figure 1: LeNet-5





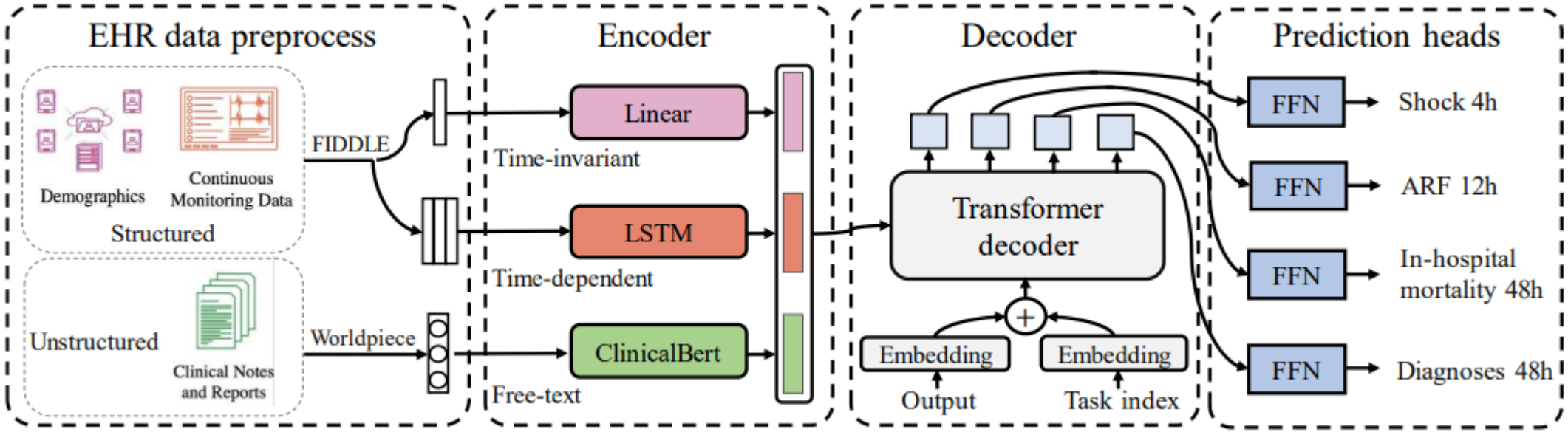
CNN

RNN

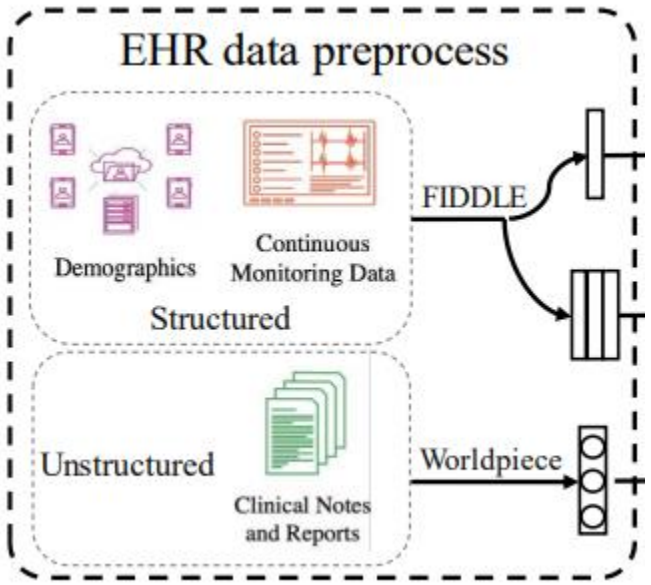
GNN

3. Objective

- **Inductive Bias**
  - But in a clinical research, shows that acute respiratory failure is related to a mortality rate of 35%–46% in clinic



4. Method



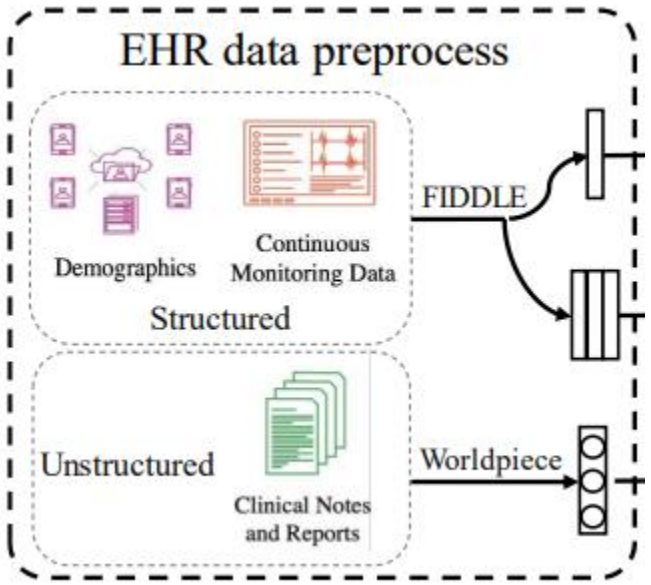
◦ **MIMIC-III**

- 공공 의료 데이터 세트로 Israel Deaconess Medical Center에서 2001부터 2012년까지 환자 데이터를 공개
- 58,976 개의 입원기록이 존재하고 Demographics, Vital signs, Laboratory measurements 등 환자에 대한 여러 정보를 확인
- 해당 연구에서는 MIMIC-III의 인구역학정보, 검사기록, Clinic note를 활용

◦ **FIDDLE**

- 기계학습 모델의 해석 가능성을 높이기 위한 방법으로, 모델의 입력과 출력 간 관계를 파악하고 모델의 결정에 영향을 미치는 요인을 파악하는 방법
- 일반적으로 LinearRegression 모델을 사용하여 입력과 출력 간의 선형 관계를 학습하고 특성 중요도를 계산

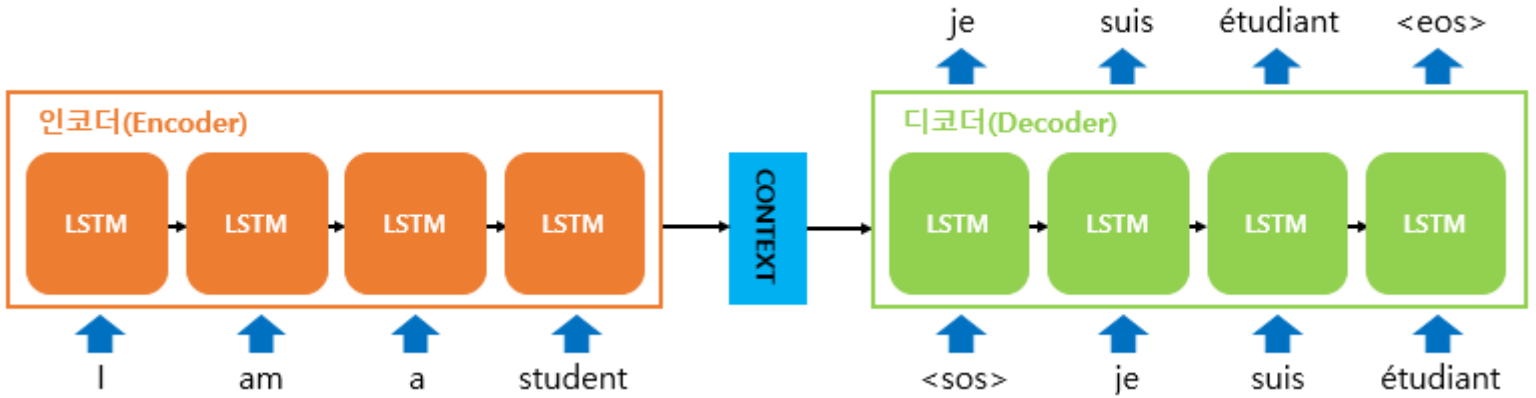
4. Method



- **Worldpiece**
  - 구글의 NMT(Neural Machine Translation)에서 제안한 워드피스(wordpiece) 토크나이저를 사용하여 기존의 단어를 더 작은 단위의 워드피스로 분리
  - 워드피스 토크나이저는 기존의 단어를 하위 단어(subword)로 나누는 작업을 수행하여, 새로운 단어에 대한 대처력을 높이는데 도움을 줌
  - Ex. 자동차 = 자동 + 차
- **Input**
  - Demographics > FIDDLE > Table
  - Continuous Monitoring Data > FIDDLE > Time Series
  - Clinical Notes and Reports > Wordpiece > Word Vector

4. Method

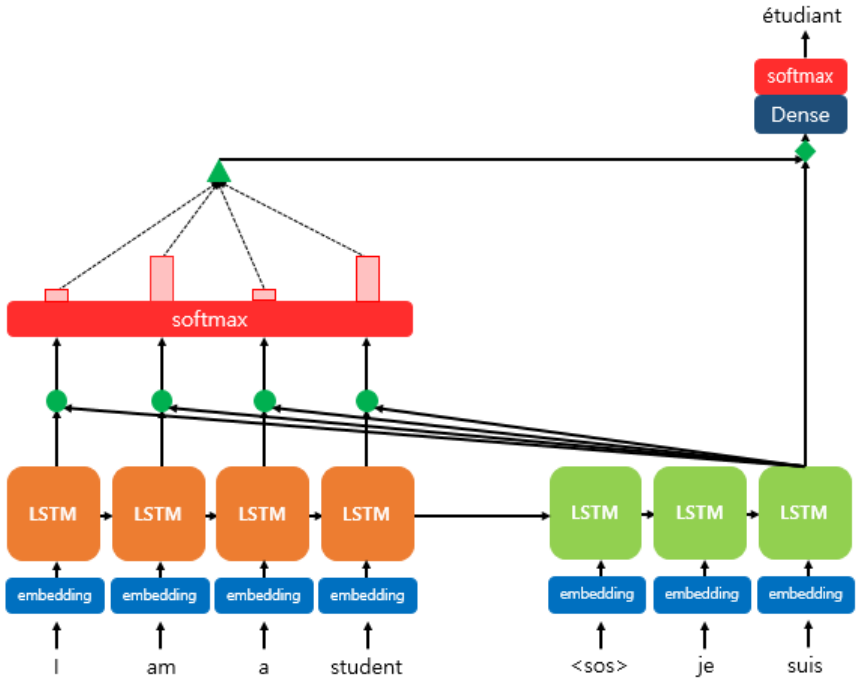
◦ **Seq2Seq**



- 인코더를 통해 Context Vector를 만들고 디코더에서 해당 Vector와 Label을 활용하여 학습을 진행
- 기계번역에서 높은 성능을 보임
- Context Vector를 만드는 과정에서 정보 손실이 발생하고, 기울기 소실 문제를 해결하지 못함

4. Method

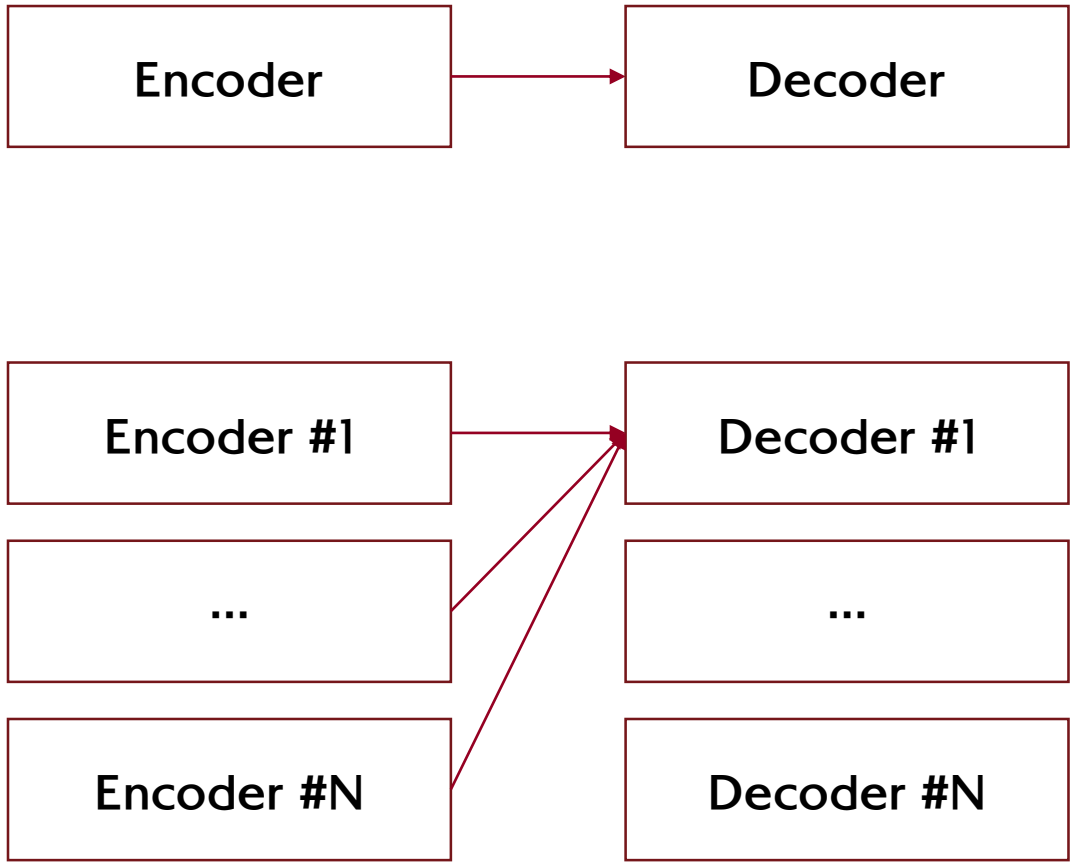
- **Attention**



- Query, Key, Value를 활용하여 매 시점마다 전체 입력 문장을 참고하여 각 시점마다 중요한 특징을 선별
- Self-attention은 Query와 Key를 같은 값에서 추출하여 미지의 Inductive Bias를 찾도록 함

4. Method

◦ Transformer



- Encoder와 Decoder를 여러 층으로 쌓아올린 구조로 각 층에서 Attention 기법을 활용
- 디코더는 이전 정보들 중 현재 예측해야 할 결과와 관련 있는 정보만 고려
- 위치 정보가 무시된다는 단점은 Positional Encoding을 통해 해결

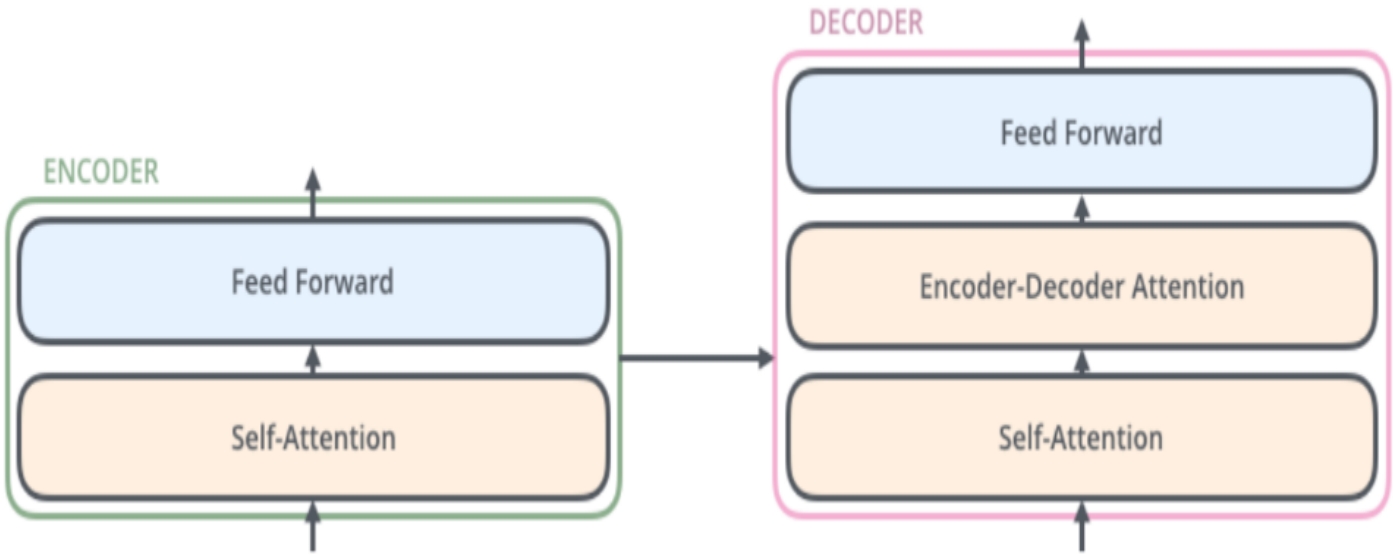
$$PE_{(pos,\,2i)} = sin(pos/10000^{2i/d_{model}})$$

$$PE_{(pos,\,2i+1)} = cos(pos/10000^{2i/d_{model}})$$

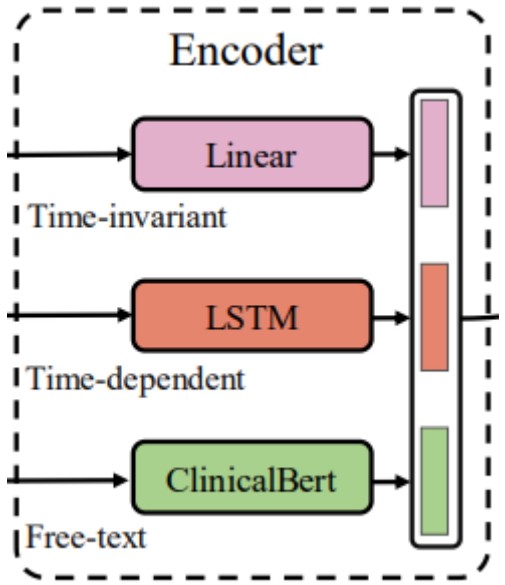
Ex. 첫 번째의 '나'와 중간의 '나'는 같은 단어이지만 다른 값을 가짐

4. Method

- **Transformer**
  - 1. Encoder Self-Attention: 입력 벡터에서 중요 정보를 추출
  - 2. Decoder Masked Self-Attention: 출력 벡터에서 중요 정보를 추출. 이때 Mask를 활용하여 결과값의 순서를 고려
  - 3. Encoder-Decoder Attention: Encoder에서 선별된 정보와 Decoder에서 선별된 정보를 바탕으로 Attention을 진행



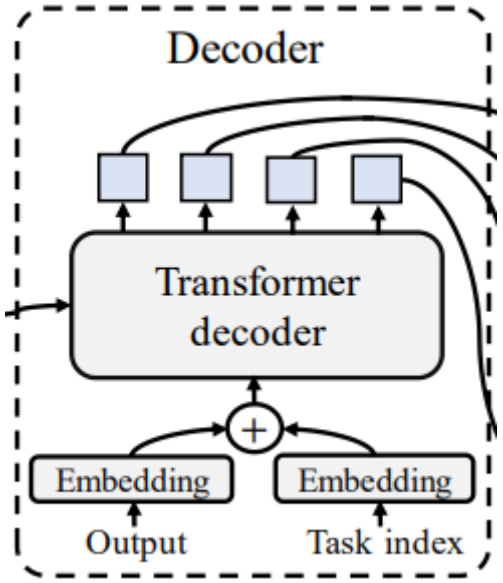
4. Method



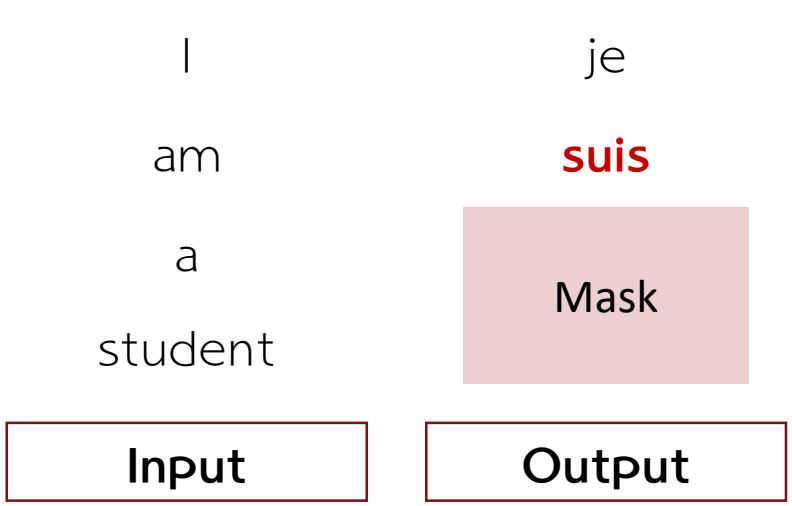
- **ClinicalBert**
  - 구글의 Bert 모델을 기반으로 의료분야에 맞게 전이학습기켜 만들어진 자연어 처리 모델
  - 여러 의학 용어를 vocabulary에 추가하고 의료분야의 데이터를 통해 fine tuning을 진행

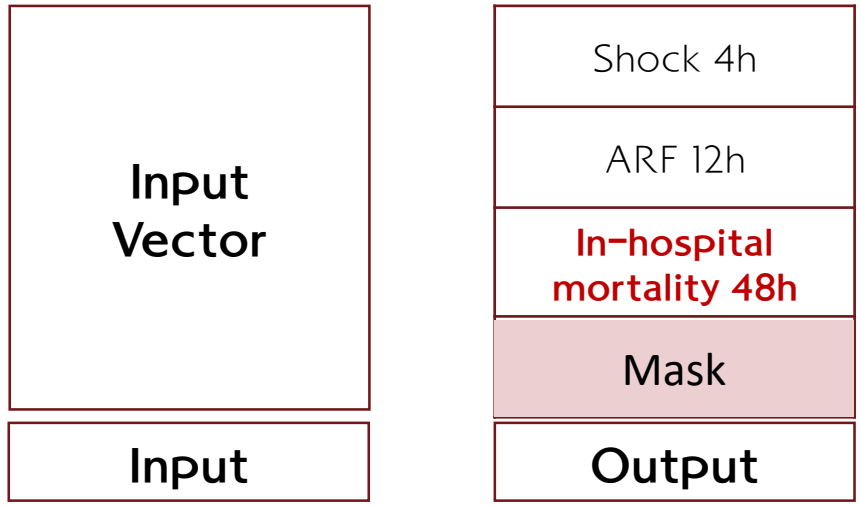
- **Encoding**
  - Table > Linear Regressor > d1의 크기를 가진 Vector A
  - Time Series > LSTM > d2의 크기를 가진 Vector B
  - Word Vector > ClinicalBert > d3의 크기를 가진 Vector C
  - Vector A, Vector B, Vector C concat > d(d1 + d2 + d3)의 크기를 가진 Vector

4. Method



- **Transformer Decoder**
  - Positional Encoding을 통해 입력과 출력의 위치 정보를 유지
  - Transformer 기법을 활용하여 모델을 구축
  - 기계번역과 비슷한 방식으로 동작





5. Result

◦ **결과**

- 단일 modal보다 multi-modal을 사용했을 때, 성능이 더 높음. 이는 다양한 자료에서 더 의미있는 모델을 만들 수 있음을 시사함
- 단일 task보다 multi-task를 수행했을 때, 성능이 더 높음. 이는 서로 연관되어 있는 task를 수행할 때 더 의미있는 모델을 만들 수 있음을 시사함

TABLE II
PERFORMANCE OF INDEPENDENT MODALITY MODELING.

Modality	ARF 4h	ARF 12h	In-hospital mortality 48h	Diagnoses 48h
	AUROC	AUROC	AUROC	Recall@10
Structured data only	0.802	0.756	0.857	0.281
Unstructured data only	0.742	0.715	0.839	0.336
Both (UniMed)	**0.836**	**0.781**	**0.892**	**0.385**

TABLE III
PERFORMANCE OF DIFFERENT DECODERS IN SINGLE-TASK AND MULTI-TASK.

Decoder	Task	Shock 4h	ARF 12h	In-hospital mortality 48h	Diagnoses 48h
	Metric	AUROC	AUROC	AUROC	Recall@10
GRU	Single-task	0.815	0.746	0.865	0.316
	Multi-task	0.821	0.760	0.871	0.331
GRU with attention	Single-task	0.823	0.758	0.867	0.329
	Multi-task	0.836	0.770	0.886	0.363
Transformer (UniMed)	Single-task	0.831	0.764	0.873	0.331
	Multi-task	**0.836**	**0.781**	**0.892**	**0.385**

5. Result

◦ Limitation

- 사용가능한 timestamp의 제한
- 낮은 질병명 예측

◦ 고찰

- EHR의 다양한 Modal을 활용하는 방식을 제시
- 상관성을 가진 task를 동시에 진행할 때, 성능이 개선될 수 있음을 제시
- 적절한 Inductive Bias가 모델의 성능에 도움이 됨. 따라서 시간이 지날수록 AI개발에 있어 **도메인 지식의 중요성이 강조**될 것임
- MIMIC-III의 데이터만을 활용하여 일반화 성능에 대한 의문점이 존재. **외부 검증**과 함께 **여러 병원의 데이터를 통합하는 연구**가 선행된다면 더 의미있는 결과를 얻을 수 있을 것임
- Multi Modal은 성능을 개선시킬 수 있지만, **모델 해석을 어렵게 만듦**. 따라서 Multi Modal로 개발된 모델을 해석할 수 있는 방안이 마련되어야 함
- **모델 성능의 개선이 예측 타겟의 시간적 상관성을 입증하는 수단은 아님**. 이를 증명할 수 있는 추가적인 방안이나 도메인 지식이 필요하며 이는 인과관계를 규명하는 단서가 될 수 있음

감사합니다