Анализ литературы

Актуальность: необходимость в алгоритме предсказания состояния пациента в будущем по предыдущим состояниям для принятия более детального решения.

Цель – рассмотреть существующие решения в сфере предсказания

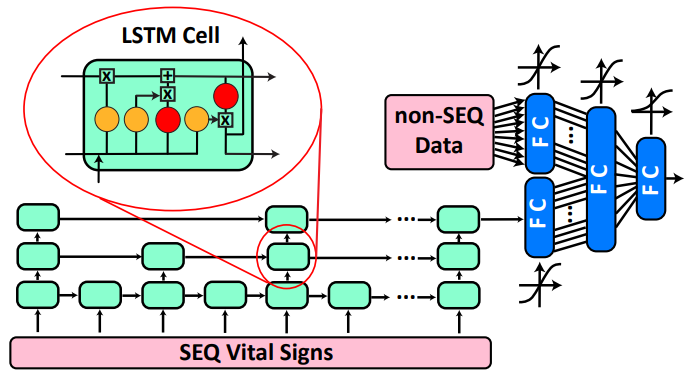
Ключевые слова для поиска – все, что связано с предсказанием нескольких непрерывных величин regression, treatment progression predict, COVID, time-series prediction, EHR preditction

1. В работе «A Comprehensive Benchmark for COVID-19 Predictive Modeling Using Electronic Health Records in Intensive Care: Choosing the Best Model for COVID-19 Prognosis» сравниваются модели разных видов – классические ML и DL. Предсказываются две величины – возможность выздоровления и количество дней до выздоровления. Лучшие методы для задачи поиска значения дней до конца лечения (по метрике MAE) – Random Forest, XGBoost, MLP, LSTM, TCN (<https://arxiv.org/abs/2209.07805>)

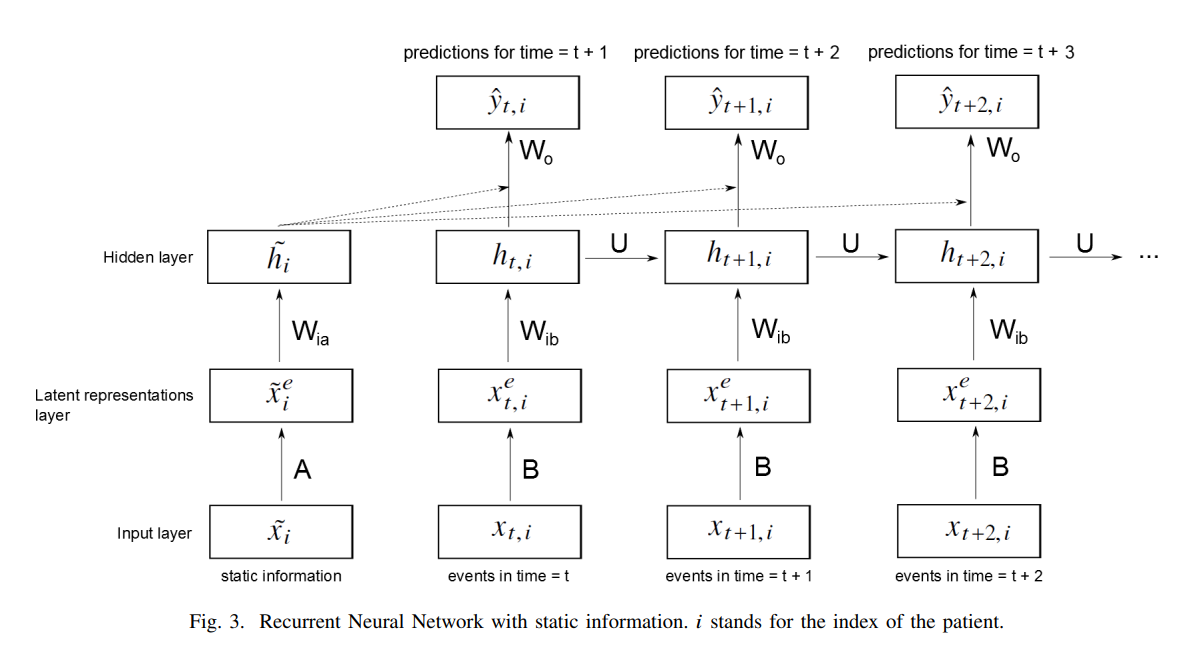
Изображение выглядит как стол

Автоматически созданное описание

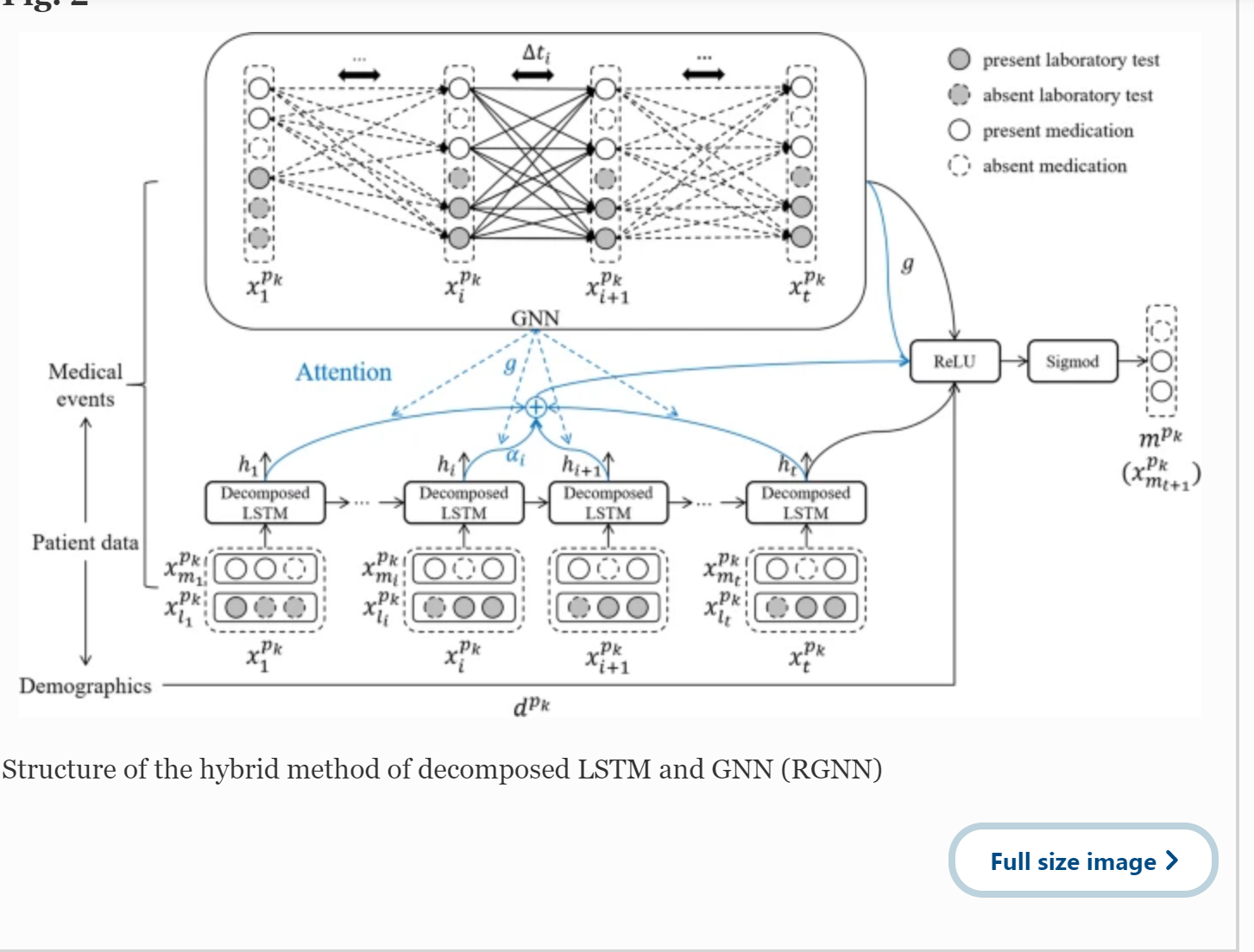
1. В статье «Deterioration Prediction using Time-Series of Three Vital Signs and Current Clinical Features Amongst COVID-19 Patients» (<https://arxiv.org/pdf/2210.05881.pdf> ) используются статичные и динамические данные для оценки вероятности осложнений от заболевания. В предложенной схеме статичные данные комбинируются с выводом LSTM в последних слоях нейросети.



1. В работе «Predicting Clinical Events by Combining Static and Dynamic Information Using Recurrent Neural Networks»C. Esteban комбинируют статичные и динамичные показатели для предсказания следующих назначений доктора. Для каждого периода времени к текущему состоянию ячейки RNN добавляют представление статичных данных о пациенте, выполненное отдельным PCA методом и на основе этого выполняют предсказание для следующего периода процесса. Так же можно сделать уже для предсказания конкретных значений(<https://arxiv.org/abs/1602.02685>)



1. В работе «Predicting the outcome for COVID-19 patients by applying time series classification to electronic health records» (Davi Silva et. Al 2022) сравнивают модели и способы их применения для предсказания исхода лечения COVID-19. В данном случае используют подход с предсказанием по целому вектору параметров и среднее по предсказаниям по отдельным значениям. В качестве модели используют свертку mini-ROCKET. Во втором случае результат применения по метрике точности – 97%. В моем случае так же можно разбить на несколько моделей для каждого параметра. (<https://link.springer.com/content/pdf/10.1186/s12911-022-01931-5.pdf> )
2. В работе A hybrid method of recurrent neural network and graph neural network for next-period prescription prediction ([Sicen Liu](Sicen%20Liu) et al.) предсказывают следующие этапы лечения с помощью комбинации LSTM и GNN сетей. Из этой схемы можно использовать схему развернутой LSTM с конкатенацией состояний ячеек (<https://link.springer.com/article/10.1007/s13042-020-01155-x> )



1. **Подходы для коротких временных рядов.** В работеShort Time Series Forecasting: Recommended Methods and Techniques (Mariel Abigail Cruz-Nájera, 2022) было показано на примере ряда уровня преступности, что классические методы, такие как SMA, ARIMA, Hybrid, and Jaganathan выигрывают на коротких рядах, в отличие от продвинутых LightGBM, ANN. (<https://www.researchgate.net/publication/361299592_Short_Time_Series_Forecasting_Recommended_Methods_and_Techniques> **)**
2. **AutoML подходы. Fedot.** Фреймворк поддерживает отдельно задачу предсказания следующих значений временного ряда [2] *(“ts\_forecasting”).* Предсказания следующих значений ряда как в статье [3] не подходит, так как у нас 1992 ряда для каждого, а не один длинный процесс. В статье [1] по предсказанию многомерных рядов несколько рядов комбинируются и по ним предсказываются значения ряда в будущем. Аналогично можно разделить на 13 рядов для каждого признака и по ним предсказывать, но опять же у нас много независимых процессов для каждого пациента.
3. **Вставка пропущенных значений**. Для добавления пропущенных значений в многомерных рядах в работеAn Effective Multivariate Time Series Pre-Train Model under Missing Data <https://arxiv.org/pdf/2209.07798.pdf> представлена модель DBT-DAE, сеть типа decoder-encoder. Иной подход представлен в работе Multiple Imputation by Chained Equations in Praxis: Guidelines and Review (<https://www.researchgate.net/publication/316789124_Multiple_Imputation_by_Chained_Equations_in_Praxis_Guidelines_and_Review>) 