

НАЗНАЧЕНИЕ

Прогноз развития сердечно-сосудистых заболеваний (ССЗ) в течение 10 последующих лет на основе методов машинного обучения. Для построения модели были взяты признаки, используемые во Фрамингемской шкале: пол, возраст, САД, холестерин, холестерин ЛПВП, прием АГТ, факт курения.

МЕТОДЫ

В качестве модели использована последовательная нейронная сеть с одним входным, двумя скрытыми и одним выходным слоем. Для предотвращения переобучения использовано исключение («dropout»). На каждом слое используется функция «dense» для полного соединения слоев друг с другом. В скрытых слоях используется функция активации «relu».

ДАННЫЕ

Для обучения модели использовались 2 типа данных (табл.1):

- 1) Данные клинического исследования [Framingham Heart Study](#). Для проведения обучения модели из вышеуказанного полного дата-сета был выбран набор данных, состоящий из 2 588 пациентов без сердечно-сосудистых заболеваний на момент обследования, из которых 606 (23.4% от когорты) получили сердечно-сосудистое заболевание в течение 10 лет с момента обследования.
- 2) РЕГИОНАЛЬНЫЙ набор данных, собранный у 438 пациентов на Северо-Западе России, который был использован для повторного обучения. Популяция исследования была разделена в наборе данных на "обучающую" выборку (80%), в которой были получены алгоритмы определения рисков и «тестовую» выборку (20%), которая применялась для тестирования и валидации алгоритмов (рис. 1).

РЕЗУЛЬТАТЫ

Показатели точности модели по ROC анализу составляют: для модели, обученной только на данных клинического исследования Accuracy: 78.8%, AUC: 0.78; для дообученной модели на данных как клинического исследования, так и регионального набора данных: Accuracy: 80.2%, AUC: 0.83 (рис. 2).

Для сравнения, аналогичные показатели для Фрамингемской шкалы имеют более низкие метрики: точность Accuracy: 70.0%, качество AUC: 0.59.

Таблица 1 - СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛИ

	PATIENTS (n)	AGE (y.o)	MALE (%)	SMOKE (%)	TOT CHOL (mg/dl)	HDLC (mg/dl)	SYSBP (mmHg)	BPME DS (%)	CVD (%)
FRAMINGHAM DataSet	2 588	49.9 ± 8.6	56	49	6.16 ± 1.16	1.26 ± 0.39	132.8 ± 22,3	55	23.4
REGIONAL DataSet	438	47,9 ± 5,7	58,9	37,7	5.16 ± 0.75	1.32 ± 0.30	128.3 ± 14.8	21.2	26.8

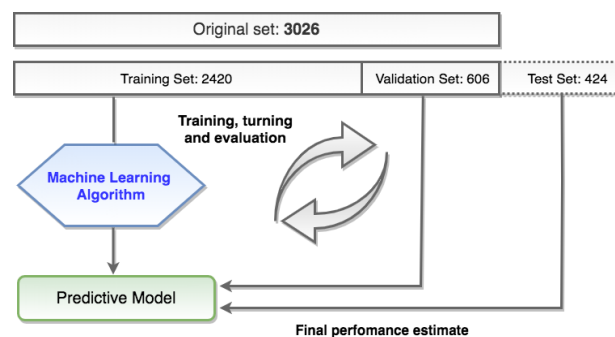


Рисунок 1 – ПРОЦЕСС ОБУЧЕНИЯ, ВАЛИДАЦИИ И ТЕСТИРОВАНИЯ МОДЕЛИ НА ДАННЫХ

ОПРЕДЕЛЕНИЕ РИСКА ПАЦИЕНТА:

- НИЗКИЙ** – минимальный риск развития ССЗ
- ВЫСОКИЙ** – значительный риск развития ССЗ

ПУБЛИКАЦИИ АВТОРОВ:

1. [European Heart Journal, Volume 40, Issue Supplement 1, October 2019](#)
2. [Врач и информационные технологии • №3 2019](#)
3. [ESC Congress 2020 The Digital Experience](#)

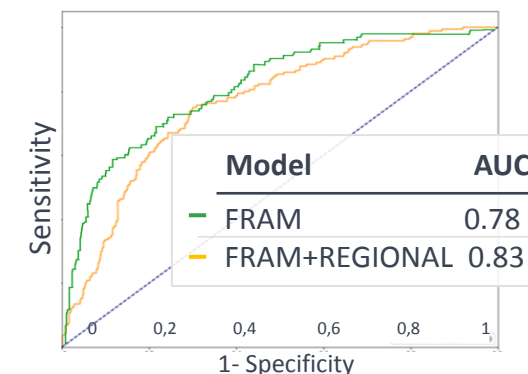


Рисунок 2 – ROC КРИВАЯ ДЛЯ МОДЕЛЕЙ, ОБУЧЕННЫХ НА ДАННЫХ:

FRAM - данные [Framingham Heart Study](#)
REGIONAL - дообучение на региональном наборе данных