

НАЗНАЧЕНИЕ

Прогноз смерти от ишемической болезни сердца и инсульта в течение 10 последующих лет на основе методов машинного обучения. Для построения модели были взяты признаки, используемые в клинической шкале SCORE: пол, возраст, САД, холестерин, факт курения, и дополненные два признака (индекс массы тела и частота сердечных сокращений).

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 4 363 пациентов без сердечно-сосудистых заболеваний на момент обследования, из которых 852 (19,5% от когорты) получили сердечно-сосудистое заболевание в течение 10 лет с момента обследования.

2) РЕГИОНАЛЬНЫЙ набор данных, собранный у 411 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=79%, AUC = 0,78; для дообученной модели на данных как клинического исследования, так и регионального набора данных: Accuracy = 79%, Se= 75%, AUC = 0.86 (рис. 2). Для шкалы Score, полученные на том же тестовом наборе данных имеют более низкие метрики: Se= 65%, AUC=0.81.

Таблица 1 - СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛИ

	PATIENTS (n)	AGE (y.o)	MALE (%)	SMOKE (%)	TOT CHOL (mg/dl)	HEART RATE (beats per minute)	SYSBP (mmHg)	BMI	CVD DEATH (%)
FRAMINGHAM DataSet	4 363	49,9±8,6	56	49	6,16±1,16	75,8±12,0	132,8±22,3	25±4,1	19,5
REGIONAL DataSet	411	47,9±5,7	58,9	37,7	5,16±0,75	57,00±21,7	128,3±14,8	27,5±4,4	21,1

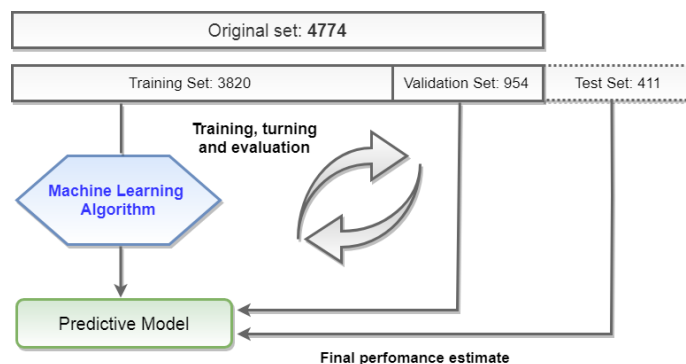


Рисунок 1 – ПРОЦЕСС ОБУЧЕНИЯ, ВАЛИДАЦИИ И ТЕСТИРОВАНИЯ МОДЕЛИ НА ДАННЫХ

ОПРЕДЕЛЕНИЕ РИСКА ПАЦИЕНТА:

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

ПУБЛИКАЦИИ АВТОРОВ:

- [European Heart Journal, Volume 40, Issue Supplement 1, October 2019](#)
- [ВРАЧ. НАУЧНО-ПРАКТИЧЕСКИЙ ЖУРНАЛ, №5 2020](#)
- [ESC Congress 2020 The Digital Experience](#)

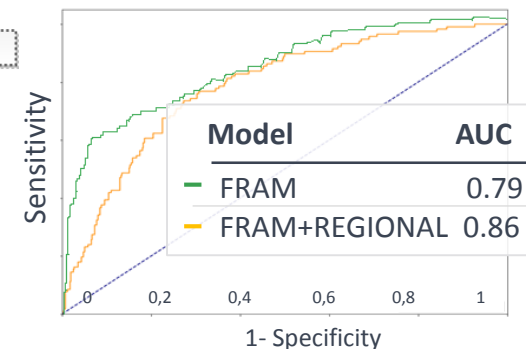


Рисунок 2 – ROC КРИВАЯ ДЛЯ МОДЕЛЕЙ, ОБУЧЕННЫХ НА ДАННЫХ:

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

СВИДЕТЕЛЬСТВА И ПАТЕНТЫ

- РУ на медицинское изделие № РЗН 2020/9958 от 03.04.2020
- Свидетельство ПрЭВМ №2020619583 от 19.08.2020