Бусыгина Юлия Сергеевна, Константиниди Виктория Константиновна, Жалялова Диана Руслановна, Лушина Ульяна Алексеевна, Дыйканова Лиля Куатбековна, Темирбеков Шихсаид Магомедкамилович, Панина Ульяна Валерьевна
Первый Московский государственный медицинский университет имени И. М. Сеченова Министерства здравоохранения Российской Федерации (Сеченовский Университет)

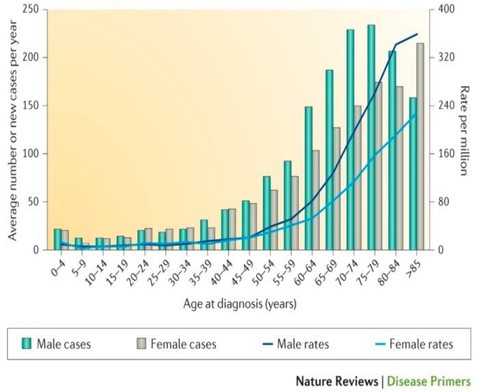
ПРОГНОЗИРОВАНИЕ ВЕРОЯТНОСТИ ЛЕТАЛЬНОГО ИСХОДА ГЕМАТОЛОГИЧЕСКИХ ПАЦИЕНТОВ ПО ДАННЫМ ЭЛЕКТРОННЫХ МЕДИЦИНСКИХ КАРТ С ПРИМЕНЕНИЕМ МАШИННОГО ОБУЧЕНИЯ

ВВЕДЕНИЕ.

Острый миелоидный лейкоз (ОМЛ)– это клональное опухолевое заболевание кроветворной ткани, связанное с мутацией в клетке-предшественнице гемопоэза, следствием которой становится блок дифференцировки и бесконтрольная пролиферация незрелых миелоидных клеток.

Эпидемиология ОМЛ:

- **Заболеваемость — 3-5 человек на 100 тыс. населения в год**
- **18% лейкозов у детей**
- **Медиана возраста диагностики составляет 68 лет**



Цель:
разработать и оценить алгоритмы машинного обучения для предсказания смертности во время госпитализации, на основе данных, доступных на момент поступления пациента с ОМЛ.

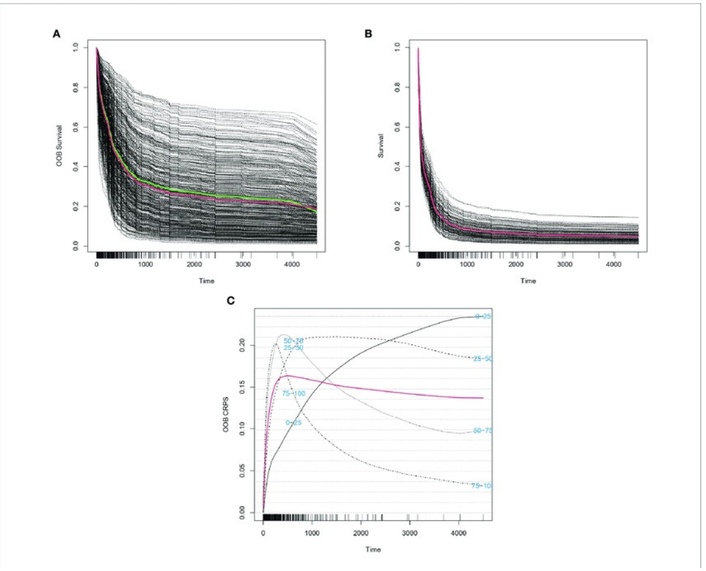
МАТЕРИАЛЫ И МЕТОДЫ.

была использована база данных из медицинских электронных карт стационарных пациентов за 2018–2019 годы, которая включала пациентов (возраст >18 лет) с диагнозом ОМЛ и получавших химиотерапию во время этой госпитализации.

Выгрузка проводилась из системы ЕМИАС ДЗМ г. Москва
Основная цель заключалась в предсказании смертности у пациентов с ОМЛ, проходящих химиотерапию, используя данные, имеющиеся перед началом химиотерапии (возраст, количество дней с момента поступления до начала химиотерапии, сопутствующие заболевания, жалобы на момент поступления).

Набор данных был разделен на обучающую (80%) и тестовую (20%) выборки.
Был проведен сравнительный анализ нескольких алгоритмов классификации: логистическая регрессия, дерево принятия решения и случайные леса.
Точность модели на тестовом наборе данных была оценена с помощью параметра «площадь под кривой» ROC (AUC, ROC).

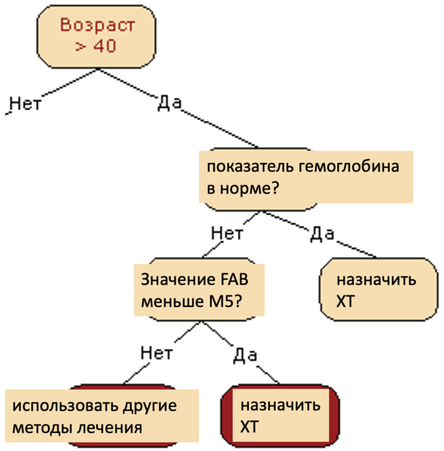
ROC-AUC кривые алгоритмов классификации: логистическая регрессия, дерево принятия решения и случайные леса:



РЕЗУЛЬТАТЫ.

В исследование были включены данные 1278 пациентов с лейкозом, медианный возраст составил 47,2 лет. Модели логистическая регрессия и случайных лесов продемонстрировали AUC=0,68, модель деревья принятия решений AUC=0,75, модель логистической регрессии AUC=0,62.

Пример работы дерева принятия решения:



Полученный результат с коэффициентами значимости факторов для прогнозирования летального исхода:

Параметр	Относительный индекс влияния на предсказание
Возраст	15.27358
ЛДГ	13.43576
FAB- классификации (**French-American-British**)	8.380744
Цитогенетическая классификация	5.140746
CD19	4.937505
TP53	4.839846
STAT5.p431	4.773225
ERk2.p	4.72874
Гемоглобин	4.143627
STAT3	3.829885
P70S6K.p	3.435698
Альбумин	3.430047

ЗАКЛЮЧЕНИЕ.

Преимущество такого метода заключается в использовании доступных параметров для предсказания смертности в стационаре пациентов с ОМЛ, которые проходят химиотерапию, а также такое решение не требует дополнительных финансовых затрат.
Данное исследование подтверждает возможность использования алгоритмов машинного обучения для создания решений в области здравоохранения