

# Объяснимый фреймворк машинного обучения для прогнозирования сердечно-сосудистых рисков с использованием структурированных медицинских данных

**Источник:** Frontiers in AI — Medicine

**Оригинал:** <https://www.frontiersin.org/articles/10.3389/frai.2026.1812622>

диагностика

кардиология

машинное обучение

объяснимый ИИ

прогнозирование рисков

## Введение

Сердечно-сосудистые заболевания (**CVD** — Cardiovascular Disease) по-прежнему остаются одной из ведущих причин смерти во всем мире. Вследствие сложности клинических данных все чаще применяются модели машинного обучения для прогнозирования риска **CVD**. Тем не менее, многие методы машинного обучения страдают от отсутствия интерпретируемости, что затрудняет их использование в клинической практике. В данном исследовании представлена интерпретируемая платформа машинного обучения для прогнозирования сердечно-сосудистого риска с использованием структурированных клинических данных.

## Методы

В данном исследовании использовался общедоступный набор данных по сердечно-сосудистым заболеваниям, состоящий примерно из 70 000 записей пациентов. Он содержит различные демографические, физиологические и связанные с образом жизни переменные, которые обычно используются при оценке сердечно-сосудистого риска. Для разработки трех моделей

машинного обучения, а именно **LogisticRegression()** (логистическая регрессия), **RandomForestClassifier()** (случайный лес) и **GradientBoostingClassifier()** (градиентный бустинг), была проведена стратифицированная перекрестная проверка (Stratified Cross Validation) по пяти блокам. Были измерены показатели эффективности моделей по различным метрикам оценки, таким как точность (**accuracy**), прецизионность (**precision**), полнота (**recall**), **F1-score** и площадь под ROC-кривой (**AUC-ROC** — Area Under the Receiver Operating Characteristic Curve). Для объяснения как глобального, так и локального вклада признаков в целях повышения интерпретируемости использовался метод **SHAP** (**Shapley Additive Explanations** — аддитивные объяснения Шепли).

## Результаты

Модели, оцененные в ходе эксперимента, продемонстрировали схожую прогностическую эффективность, при этом ансамблевые методы показали лучшие результаты. Модель **Voting Ensemble** (голосовой ансамбль) заняла второе место с показателем **AUC** 0,793 (**Gradient Boosting** показал наивысшую прогностическую эффективность: 0,794). Модели достигли показателя **AUC**, значительно превышающего базовую модель логистической регрессии, которая показала результат 0,773. Более высокая точность ансамблевых моделей обусловлена главным образом их способностью улавливать нелинейные взаимодействия между признаками в наборе данных.

## Обсуждение

Что касается наиболее влиятельных предикторов во всех моделях, анализ объяснимости показал, что преобладающими были **возраст, артериальное давление, уровень холестерина** и **вес**. Благодаря применению методов объяснимого искусственного интеллекта в сочетании с моделями машинного обучения, эти результаты демонстрируют, как подобные подходы могут привести к более прозрачному и интерпретируемому прогнозированию сердечно-сосудистого риска. Данная платформа демонстрирует потенциал объяснимого машинного обучения в содействии принятию клинических решений и укреплению доверия к прогностическим моделям в здравоохранении.

---

---

Перевод выполнен: 15.05.2026 | ai4med.ru

Машинный перевод. Рекомендуем сверять с оригиналом при клиническом использовании.