

УДК 004.89

ПРОБЛЕМЫ ОБЕСПЕЧЕНИЯ КАЧЕСТВА ЭЛЕКТРОКАРДИОГРАФИЧЕСКИХ ДАННЫХ ПРИ ОБУЧЕНИИ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Д. В. ПАНЬКИН

Научный руководитель В. В. ДУБРОВИН, канд. техн. наук, доц.
Тамбовский государственный технический университет
Тамбов, Россия

Введение. Электрокардиографические сигналы являются одним из наиболее информативных источников данных для диагностики сердечно-сосудистых заболеваний. Однако их низкая амплитуда и высокая чувствительность к внешним воздействиям делают их крайне уязвимыми для помех. Машинное обучение предъявляет высокие требования к качеству входных данных, поскольку шумные сигналы приводят к снижению точности классификации и потере диагностически значимой информации. При этом применение комплексной гибридной предобработки позволяет повысить точность классификации по сравнению с моделями, обученными на неочищенных данных. Поэтому разработка эффективных пайплайнов подготовки данных является критически важным этапом создания систем автоматической диагностики [1].

Основной часть. Наиболее эффективным подходом к подготовке ЭКГ-данных признаётся многоэтапная каскадная гибридная фильтрация. Первый этап заключается в удалении дрейфа изолинии. Стандартным подходом является применение фильтра верхних частот с частотой среза около 0,5 Гц. В некоторых исследованиях также используется вейвлет-фильтрация с вейвлетами, позволяющая отделить низкочастотные компоненты без значительного искажения формы QRS-комплекса. ЭКГ-сигналы очищаются от основных видов помех, включая дрейф изолинии, сетевую наводку частотой 50...60 Гц, мышечные шумы и артефакты движения. Для этого последовательно применяются частотные методы фильтрации: высокочастотная фильтрация для устранения низкочастотного дрейфа, режекторная фильтрация для подавления сетевой наводки и низкочастотная либо адаптивная фильтрация для снижения мышечного шума.

На следующем этапе выполняется контроль качества очищенного сигнала, выявляются аномальные значения и повреждённые участки с использованием статистических методов, в частности Z-score и межквартильного размаха (IQR). При обнаружении пропусков или локально повреждённых сегментов производится их восстановление посредством интерполяции. После коррекции сигналов выполняются нормализация амплитуды, обеспечивающая сопоставимость ЭКГ-записей, и сегментация сигнала на отдельные участки с использованием скользящего окна.

В результате формируется очищенный, нормализованный и структурированный набор ЭКГ-данных, пригодный для последующего обучения моделей

машинного обучения. Предложенный пайплайн позволяет уменьшить влияние шумов и выбросов при сохранении диагностически значимых характеристик ЭКГ-сигнала, что создает предпосылки для повышения качества классификации сердечно-сосудистых состояний [2].

На рис. 1 представлена обобщенная схема конвейера подготовки данных для обучения моделей машинного обучения на основе ЭКГ-сигналов.

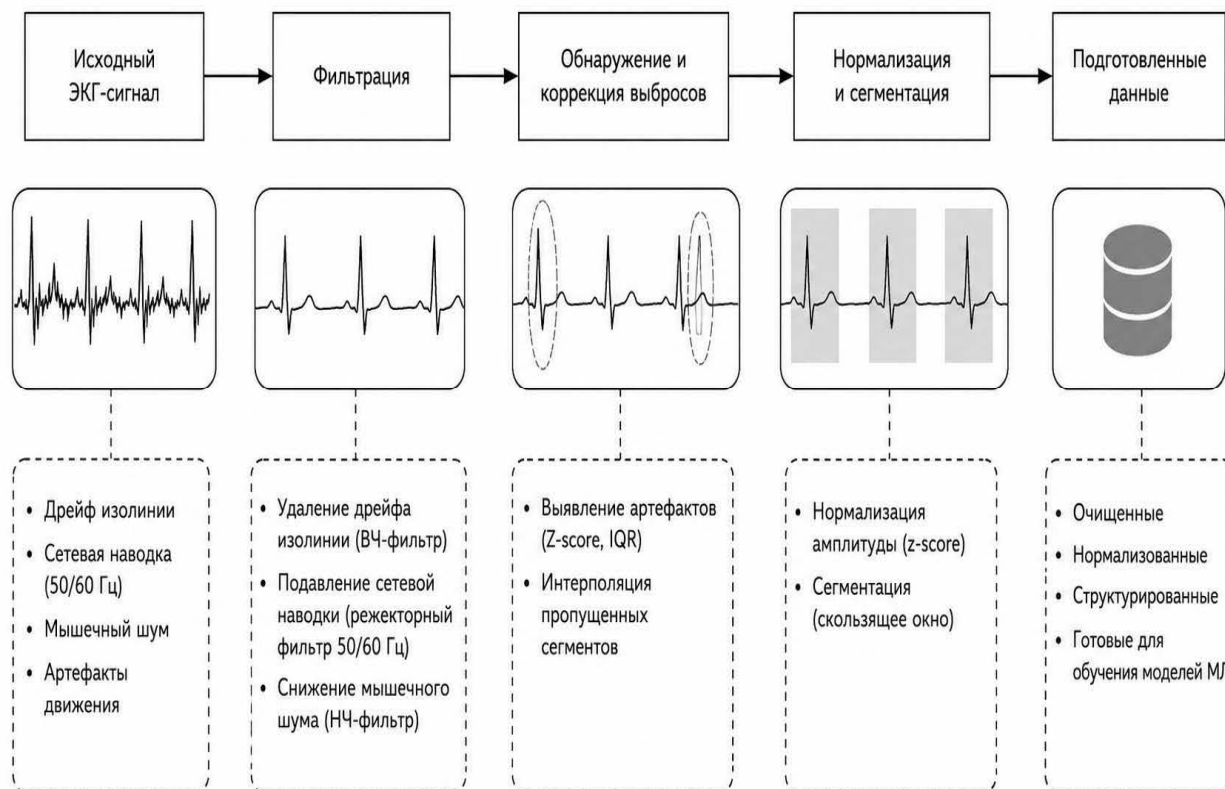


Рис. 1. Схема пайплайна предварительной обработки ЭКГ-сигналов для подготовки данных к обучению моделей машинного обучения

Заключение. Подготовка и очистка данных являются критически важным этапом построения систем машинного обучения для диагностики сердечно-сосудистых заболеваний по ЭКГ. Основная проблема заключается в многообразии типов шумов, воздействующих на сигнал, что требует их последовательного устранения без потери диагностически значимой информации.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Интеграция технологий искусственного интеллекта в анализ электрокардиограммы: современные тенденции и перспективы: обзор / Е. В. Хазова, П. А. Шулаева, Р. Р. Савзиханова, М. И. Малкова // Российский кардиологический журнал. – 2026. – Т. 31, № 25. – С. 6835.
2. Noise-Resilient Data Normalization and Outlier Mitigation Using Hybrid Filtering for Enhanced Cardiovascular Disease Prediction. IEEE Xplore, 2026.