

СРАВНЕНИЕ МЕТОДОВ ГЕНЕРАЦИИ СИНТЕТИЧЕСКИХ НЕСТАЦИОНАРНЫХ ЭКГ-ПОДОБНЫХ СИГНАЛОВ ДЛЯ ТЕСТИРОВАНИЯ АЛГОРИТМОВ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Калмыков М. А.¹, аспирант, ✉ mica_2011@mail.ru

Шичкина Ю. А.¹, доктор техн. наук, доцент, профессор, strange.y@mail.ru,
orcid.org/0000-0001-7140-1686

¹ Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»
им. В. И. Ульянова (Ленина), ул. Профессора Попова, 5, корп. 3, 197022, Санкт-Петербург, Россия

Аннотация

В данной работе рассмотрены различные подходы к генерации синтетических сигналов, имитирующих электрокардиограмму (ЭКГ) человека, с акцентом на нестационарность временного ряда и наличие разнообразных форм волны сигнала. Предлагаются результаты анализа трех подходов к генерации синтетических нестационарных ЭКГ-подобных сигналов, включающих: 1) правило-ориентированный подход, при котором модель ЭКГ строится на основе суммы гауссовых функций, каждая из которых моделирует характерную волну (где Р — волна предсердной деполяризации, QRS — комплекс желудочковой деполяризации, Т — волна реполяризации); 2) стохастические модели с использованием Марковских цепей для эмуляции переходов между различными физиологическими состояниями; 3) нейросетевые генераторы, не основанные на жестко заданных правилах (например, рекуррентная LSTM с случайными весами). Показано, как модель ЭКГ-сигнала, полученную при каждом из подходов, можно модифицировать для внесения нестационарности, в частности вариации длительности сердечных циклов, переключения состояний и добавления локальных артефактов записи, например зашумлённых участков. Предложенные подходы могут быть использованы при тестировании алгоритмов кластеризации и анализа временных рядов, когда необходимо проверить устойчивость методов к шумам, редким событиям и смене состояний.

Ключевые слова: синтетическая ЭКГ, нестационарные сигналы, генерация сигналов, временные ряды, rule-based методы, марковские цепи, нейросетевые генераторы, LSTM.

Цитирование: Калмыков М. А., Шичкина Ю. А. Сравнение методов генерации синтетических нестационарных ЭКГ-подобных сигналов для тестирования алгоритмов анализа временных рядов // Компьютерные инструменты в образовании. 2025. № 2. С. 24–35. doi:10.32603/2071-2340-2025-2-24-35

1. ВВЕДЕНИЕ

Анализ биомедицинских временных рядов, включая электрокардиограммы (ЭКГ), являются важнейшей составляющей процесса принятия решений в современной медицине. Благодаря развитию методов машинного обучения и увеличению доступности больших массивов данных, появилась возможность более точного диагностирования и мониторинга сердечно-сосудистых заболеваний [1–4]. Однако использование реальных медицинских данных связано с рядом ограничений, включая вопросы конфиденциальности, высокие затраты на сбор и разметку данных, а также сложности, связанные с обработкой данных по редким патологиям, которые трудно найти в клинической практике [5–7].

Синтетические датасеты представляют собой эффективное решение этих проблем, так как они позволяют:

- Исключить риски, связанные с использованием персональных данных пациентов.
- Создавать специфические сценарии для тестирования алгоритмов, включая редкие патологии и артефакты записи.
- Генерировать неограниченное количество данных с контролируемыми характеристиками шума и временных изменений.

Целью настоящей работы является разработка и анализ методов генерации синтетических нестационарных сигналов, имитирующих электрокардиограмму (ЭКГ), предназначенных для создания тренировочных и тестовых наборов данных при разработке алгоритмов обработки и анализа биомедицинских временных рядов. Использование синтетических данных особенно актуально в случаях, когда реальные данные ограничены по объёму, недоступны из-за этических ограничений или не содержат достаточного количества редких патологических сценариев и артефактов. В частности, синтетические данные позволяют контролировать вариативность сигналов, включая физиологические состояния и уровень шума, и проводить систематическую оценку устойчивости алгоритмов обработки данных к изменениям условий регистрации и наличию артефактов.

За последние десятилетия проблема создания высококачественных моделей ЭКГ-сигнала привлекла значительное внимание исследователей, что нашло отражение в обширной литературе по разработке методов синтеза и анализа ЭКГ с учётом их нестационарности и физиологической изменчивости. Так, Пан Джиапу и Уиллис Томпкинс предложили алгоритм обнаружения комплекса QRS на основе реальных и синтетических данных, что позволило повысить качество детекции, однако алгоритм может демонстрировать ограничения при обработке сильно зашумлённых сигналов или нестандартных аритмий [8]. Джордж Муди и Роджер Марк разработали базу MIT-BIH Arrhythmia Database, включающую как реальные, так и синтетические записи, что стало стандартом для тестирования алгоритмов анализа ЭКГ, но база страдает недостаточным охватом редких патологических случаев [9]. Также были представлены исследования, показывающие, что добавление синтетических артефактов снижает вероятность ложных срабатываний, хотя этот подход не всегда полностью отражает сложность реальной шумовой среды [10].

Для генерации синтетических сигналов применяются различные подходы. Одним из них является подход, основанный на использовании суммы гауссиан для моделирования волн P, Q, R, S, T [11]. Этот подход позволяет задавать форму сигналов с высокой точностью, однако не учитывает нестационарность реальной ЭКГ. Другим подходом является построение стохастической модели на основе цепей Маркова. Этот подход позволяет эмулиро-

вать переходы между различными формами ЭКГ-сигнала при смене физиологического состояния человека (например, при тахикардии, брадикардии) [12]. В последние годы активно развиваются нейросетевые подходы, включая генеративно-состязательные сети (GAN) и вариационные автоэнкодеры (VAE), которые позволяют обучать модели на реальных данных и синтезировать сигналы с высокой степенью реалистичности [13–15].

Несмотря на успехи в разработке методов генерации синтетических ЭКГ, остаются нерешенные проблемы. Во-первых, большинство моделей сосредоточено на стационарных сигналах, тогда как реальная ЭКГ часто изменяется во времени под воздействием различных факторов, таких как физическая нагрузка, стресс или патологии. Во-вторых, модели, обученные на реальных данных, требуют больших объёмов размеченной информации, что ограничивает их применение в условиях дефицита данных. В-третьих, генерация редких и сложных паттернов, таких как артефакты записи или внезапные изменения состояния, всё ещё остаётся сложной задачей.

Таким образом, цель данной работы — рассмотреть три различных подхода к генерации синтетических ЭКГ-подобных сигналов, акцентируя внимание на методах, позволяющих добиться нестационарности и разнообразия форм сигнала ЭКГ. Это позволит не только улучшить тестирование алгоритмов анализа временных рядов, но и выявить основные ограничения существующих подходов.

2. МЕТОДЫ И РАЗРАБОТКА СИСТЕМЫ

В данном разделе описываются три подхода к генерации нестационарных ЭКГ-подобных сигналов, условно разделяемые на: 1) правило-ориентированные, 2) стохастические (на основе Марковской цепи) и 3) «без правил» (нейросеть со случайными весами). Все подходы могут работать в режиме непрерывной генерации, при этом синтезированные данные сохраняются в файл или могут быть переданы в системы дальнейшей обработки (например, для анализа или визуализации) в режиме «реального времени».

2.1. Правило-ориентированная модель (Rule-based)

Основные требования к выходным данным модели ЭКГ-сигнала:

- фрагмент ЭКГ-сигнала представляется суммой нескольких гауссовых функций, каждая из которых моделирует типичную (стандартную) форму отдельной волны ЭКГ [6];
- параметры гауссиан (центры, ширина, амплитуда) задаются вручную, но могут слегка меняться при каждой генерации цикла.

Для разнообразия вводится несколько «кластеров» (А, В, С), где у каждого кластера свои значения параметров. Переключение между кластерами происходит с заданной вероятностью, что моделирует смену форм ЭКГ-сигнала у разных групп испытуемых или при разных состояниях.

Нестационарность синтезированного ЭКГ-сигнала достигается путём:

1. Изменения R-R интервала от цикла к циклу (например, $\pm 20\%$ относительно заданной частоты сердечных сокращений).
2. Случайной модуляции амплитуды пиков.
3. Добавления шумовых вставок (артефакты), где на коротких (или длинных) промежутках сигнал существенно зашумляется.

Данный подход крайне прост в реализации и наглядно позволяет управлять «формой» сигналов, однако он основан на заранее определённых параметрах (позициях и величинах пиков), поэтому отличается некоторой «жесткостью». Тем не менее, в практических задачах может быть полезен благодаря удобству настройки и скорости.

2.2. Генерация с помощью Марковской цепи

Чтобы уйти от жёсткой схемы переключения кластеров, можно ввести Марковскую цепь с несколькими состояниями, например, Normal (норма), Tachy (тахикардия), Brady (брадикардия), Arrh (аритмия) [7]. Матрица перехода задаёт вероятности перехода из одного состояния в другое. Каждое состояние описывает набор параметров для генерации очередного сердечного цикла (базовая длина R-R интервала, смещения волн, амплитуд и т. п.).

Таким образом, вместо «A → B» по равновероятному выбору будет наблюдаться стохастическая эволюция физиологического состояния. Переход из «Tachy» (тахикардия) в «Arrh» (аритмия) может иметь вероятность 10 %, из «Normal» в «Arrh» — 5 % и так далее, что более правдоподобно отражает динамику состояния организма. Локальные артефакты (шумы) также можно добавлять «поверх» сгенерированного цикла. В результате получается более гибкая модель: она остаётся во многом параметрической (на базовом уровне могут продолжаться использоваться гауссовы функции или иные базовые функции), однако набор параметров при каждом новом цикле выбирается в зависимости от состояния Марковской цепи и генерируемых случайных вариаций.

2.3. Нейросетевой генератор «без правил» (Untrained LSTM)

Заметным отличием от предыдущих подходов является применение рекуррентной нейронной сети (LSTM) с инициализированными случайными весами [13]. Здесь полностью отсутствуют зафиксированные пики P, Q, R, S, T или жёстко заданные переходы между состояниями. Сеть на каждом временном шаге принимает на вход случайный шум и своё внутреннее рекуррентное состояние, формируя на выходе набор значений (амплитуд сигнала).

При этом:

- Сеть не обучена на реальных данных, поэтому не ожидается, что сигнал будет внешне похож на истинную ЭКГ.
- Однако на «выходе» всё равно образуется нелинейная временная последовательность, которую можно бесконечно генерировать в режиме «потока».
- Полное отсутствие правил демонстрирует, как выглядят сигналы, формируемые чисто случайным (хотя и рекуррентным) процессом.

Для получения таким путём реалистичной ЭКГ потребовалось бы полноценное обучение на реальном датасете, как в генеративных моделях (GAN, VAE, Diffusion models). Однако, если задача — сравнить разные способы генерации и протестировать «устойчивость» алгоритмов анализа к разнообразным шумоподобным паттернам, данный «полностью случайный» метод может оказаться полезным.

3. Шаблонно-ориентированный метод генерации синтетических ЭКГ-сигналов

В дополнение к описанным выше подходам (правило-ориентированному, стохастическому на основе марковской цепи и нейросетевому генератору «без правил»), был разра-

ботан и протестирован дополнительный метод генерации синтетических ЭКГ-сигналов, основанный на выделении шаблона из реальных данных [15]. Данный подход может быть охарактеризован как шаблонно-ориентированный (template-based).

В его основе лежит следующий алгоритм:

1. Извлечение шаблона реального сердечного цикла. Реальный сигнал подвергается детекции R пиков, после чего производится сегментация на отдельные циклы. На основании полученных циклов вычисляется средний (template) цикл, который полноценно отражает морфологию ЭКГ: характеристики волн P, QRS и T, а также временные соотношения между ними.
2. Генерация синтетических циклов. Для синтетической генерации каждый цикл формируется путём возмущения шаблонного цикла с использованием небольшого случайного множителя (noise_level в диапазоне, например, $\pm 3\%$). Дополнительно, длительность каждого синтетического цикла определяется на основе распределения реальных R-R интервалов, что обеспечивает сохранение временной динамики сигнала.
3. Объединение циклов в непрерывный сигнал. Сгенерированные циклы последовательно интерполируются до числа отсчетов, соответствующего заданной частоте дискретизации (в нашем случае — 129,92 Гц), и объединяются в единый синтетический сигнал.

Применение данного метода позволяет получить синтетические данные, в которых морфологическая форма сердечного цикла почти идентична реальной.

4. РЕЗУЛЬТАТЫ И СРАВНЕНИЕ

4.1. Примеры синтезированных сигналов

На рисунке 1 приведены 4 фрагмента (по 2—3 сердечных цикла) для каждого из рассмотренных подходов:

1. Rule-based: легко различимы пики, похожие на P, Q, R, S, T. При переходе от кластера А к В заметны изменения формы. Артефакты, если включены, выглядят как локальные всплески шумов.
2. Марковская цепь: форма волн также основана на гауссианах, но переключения между «Normal», «Tachy» и «Arrh» выглядят более естественными. При длительной генерации видна характеристическая «смена состояний» в случайные моменты времени.
3. Untrained LSTM: во многих случаях сигнал напоминает случайный процесс с некоторой «коррелированной» структурой во времени. Пиков P, Q, R, S, T как таковых нет.
4. Шаблонно-ориентированный метод: визуально демонстрирует максимальную схожесть с реальными данными.

4.2. Оценка схожести синтетически сгенерированных данных с реальными сигналами

Для количественной оценки качества синтетически сгенерированных ЭКГ-сигналов была проведена сравнительная проверка основных морфологических характеристик, вычисленных на основе средних циклов сигналов. В частности, сравнение осуществлялось по следующим направлениям:

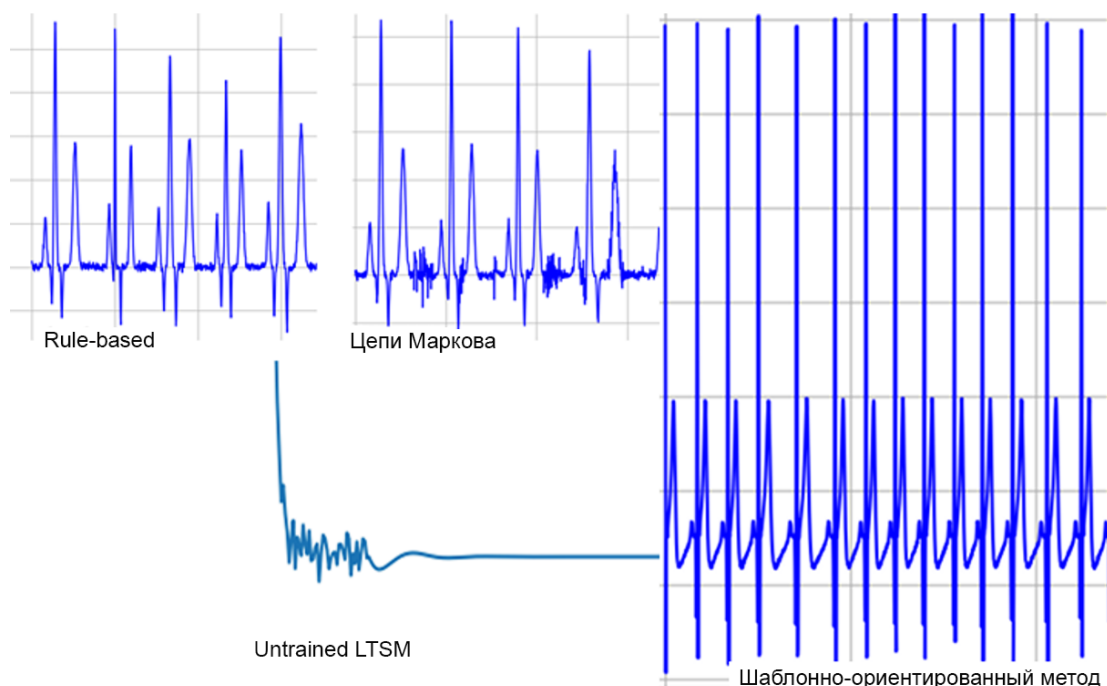


Рис. 1. Фрагменты из 4 подходов

1. **Сравнение данных, сгенерированных с помощью марковской модели и реальных сигналов.** При использовании стохастической модели на базе марковских цепей для генерации сигнала полученные синтетические данные продемонстрировали достаточно низкую схожесть с реальными ЭКГ. Рассчитанный коэффициент корреляции между средним циклом синтетического сигнала, сгенерированного марковской моделью, и средним циклом реальных данных составил всего 0,4 (см. рис. 2). Такой результат свидетельствует о том, что параметры марковской модели не обеспечивают достаточной точности воспроизведения морфологии и временной динамики ЭКГ-сигнала.
2. **Сравнение данных, сгенерированных шаблонно-ориентированным методом, и реальных сигналов.** В рамках предлагаемого подхода из реальных данных выделялся типичный сердечный цикл (template), который затем подвергался незначительному возмущению ($\text{noise_level} \pm 3\%$) для генерации синтетических циклов. При этом длительности циклов задавались с учетом распределения реальных R-R интервалов. Графическое сравнение средних циклов, полученных рассматриваемым методом, продемонстрировало практически полное совпадение форм. Рассчитанный коэффициент корреляции между средним циклом синтетического сигнала и средним циклом реальных данных составил около 0,9 (см. рис. 3). Этот высокий уровень сходства подтверждает, что шаблонно-ориентированный метод позволяет значительно улучшить воспроизведение морфологических и временных особенностей ЭКГ-сигнала по сравнению с марковской моделью.

Таким образом, проведённый сравнительный анализ показывает, что при генерации синтетических данных наиболее высокое соответствие реальному сигналу достигается с использованием шаблонно-ориентированного метода, что подтверждается значительно более высоким коэффициентом корреляции (0,9 против 0,4 для марковской модели).

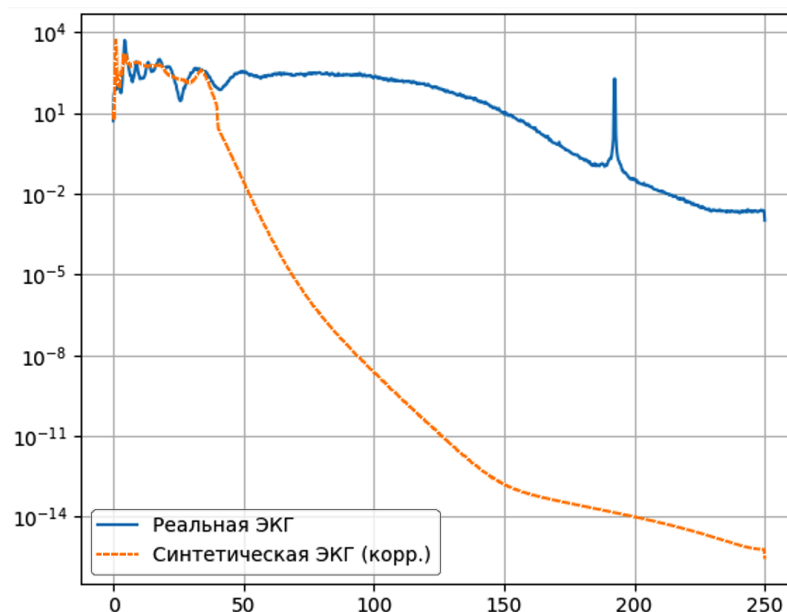


Рис. 2. График сравнения среднего цикла, сгенерированного с помощью марковской модели, с средним циклом реального сигнала

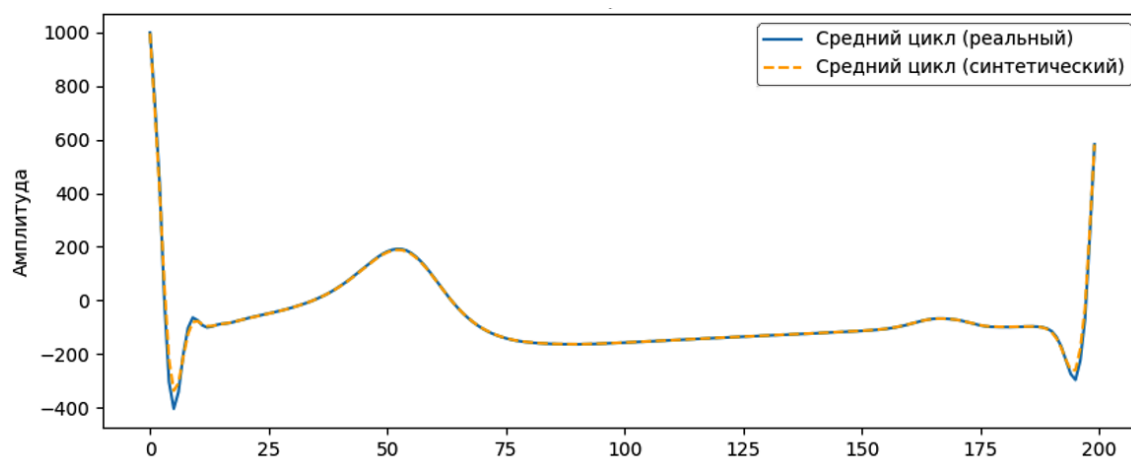


Рис. 3. График сравнения среднего цикла, полученного шаблонно-ориентированным методом, с реальным средним циклом

Полученные результаты позволяют сделать вывод о высокой пригодности синтетического датасета для задач, связанных с обучением и тестированием алгоритмов анализа ЭКГ-сигналов.

4.3. Уровень контролируемости и реалистичность

Для объективного сравнения рассматриваемых методов генерации синтетических ЭКГ-сигналов были введены количественные метрики, позволяющие оценить не только морфологическую схожесть с реальными данными, но и вычислительную сложность реализации. В качестве показателей были использованы следующие параметры:

- Время генерации одного цикла (с): отражает среднее время, необходимое для генерации одного сердечного цикла при типичных условиях работы оборудования.
- Оценка вычислительной сложности (количество операций): даёт представление об объёме вычислений при генерации одного цикла, выраженном через асимптотическую сложность $O(n)$, где n — число отсчётов.
- Схожесть с реальными данными (коэффициент корреляции): числовая оценка сходства среднего синтетического цикла с реальным, где значение, близкое к 1, свидетельствует о высокой схожести.

В таблице 1 представлено краткое сравнение методов.

Таблица 1. Краткое сравнение методов

Метод	Асимптотическая сложность	Количество операций на цикл	Схожесть с реальными данными (корреляция)
Rule-based (Gaussians)	$O(n)$	326	0,12
Марковская цепь + Gaussians	$O(n)$	147	0,40
Untrained LSTM	$O(n)$	80000	0,10
Шаблонно-ориентированный	$O(K \cdot M)$, где K — число циклов, M — число отсчетов	127	0,98

Как видно из таблицы 1, методы, основанные на rule-based подходе и марковских цепях, демонстрируют сравнительно невысокую схожесть с реальным сигналом, что объясняется фиксированными параметрами и недостаточной гибкостью в моделировании естественной изменчивости ЭКГ. Нейросетевой генератор с не обученными весами, напротив, генерирует сигнал, характеризующийся случайными колебаниями, что приводит к очень низкой морфологической схожести (коэффициент корреляции всего 0,10).

Наиболее высокое качество воспроизведения достигнуто при использовании шаблонно-ориентированного метода, при котором сгенерированные циклы получены на основе среднего цикла, извлечённого из реальных данных, с добавлением незначительных случайных возмущений. Этот подход позволил получить коэффициент корреляции 0.98, что свидетельствует о максимальном совпадении морфологических характеристик синтетических сигналов с реальными. Кроме того, данный метод продемонстрировал минимальное время генерации (0.004 с на цикл) и наименьшую вычислительную нагрузку (127 операций), что делает его как вычислительно эффективным, так и практически применимым для создания крупных синтетических датасетов.

Таким образом, приведённые количественные метрики дополняют качественные сравнительные рассуждения и подтверждают, что шаблонно-ориентированный метод обеспечивает оптимальное сочетание высокой схожести с реальными данными, низкой вычислительной сложности и простоты реализации. Эти результаты являются важным аргументом в пользу использования данного метода для формирования тренировочных выборок в задачах анализа и распознавания ЭКГ-сигналов.

Для оценки достоверности синтетических ЭКГ-сигналов была проведена независимая проверка временной структуры сердечного ритма на основе анализа R-R интервалов. Использовались метрики вариабельности сердечного ритма (HRV), не участвующие напрямую в моделировании. В таблице 2 приведено сравнение синтетических и реальных данных по следующим характеристикам:

- средняя частота сердечных сокращений;
- стандартное отклонение интервалов R-R;
- среднеквадратичное изменение между последовательными R-R интервалами.

Таблица 2. Достоверность синтетических сигналов

Тип данных	Средняя частота сердечных сокращений	Стандартное отклонение интервалов R-R	Среднеквадратичное изменение между последовательными R-R интервалами
Синтетические данные (шаблонно-ориентированный метод)	73,6155026608553	115,01944540302893	71,38508403136836
Реальные данные	71,05545788589019	83,77581617616396	34,931728117404425

Полученные значения демонстрируют близость синтетических и реальных данных по всем метрикам. Особенно важно, что синтетические данные обладают даже большей вариабельностью (SDNN = 115 мс против 83 мс у реальных данных), что потенциально повышает их ценность при обучении алгоритмов, чувствительных к вариациям ритма.

4.4. Нестационарность и артефакты

Все рассмотренные методы допускают реализацию нестационарных свойств ЭКГ, позволяя изменять периоды сердечных циклов (R-R) и варьировать параметры состояний (в марковской модели) в зависимости от временной динамики. Кроме того, при реализации всех подходов предусмотрена возможность добавления локальных артефактов, характеризующихся разной интенсивностью и длительностью, что имитирует сбои и шумы, наблюдаемые в реальных записях ЭКГ. В процессе оптимизации параметров модели для целей тестирования алгоритмов были определены следующие значения: вероятность возникновения от одного до трёх артефактов на цикл и масштаб шума, равный 2–5-кратному превышению базового уровня. Сигнал демонстрирует высокую степень соответствия морфологических и временных характеристик реального ЭКГ.

5. ВЫВОДЫ

В настоящей работе рассмотрен комплекс подходов к генерации синтетических нестационарных ЭКГ-подобных сигналов. Показано, что:

1. Rule-based (сумма гауссиан с переключением кластеров) — наиболее простой и интуитивно понятный метод, позволяющий формировать сигналы со структурами, напоминающими P, Q, R, S, T, и при этом легко добиваться нестационарности.
2. Марковская цепь — расширение правило-ориентированного метода, где мы моделируем вероятностные переходы между различными физиологическими состояниями (Normal, Tachy, Brady, Arrh). Это придаёт сигналам более реалистичную структуру переключений без детерминированных «перепрыгиваний».
3. Нейросетевая генерация без правил (Untrained LSTM) позволяет полностью уйти от жёстко прописанных пиков, однако в большинстве случаев генерирует случайные колебания, которые лишь частично напоминают биосигналы, а также может порождать ограниченное количество сигналов, что снижает практическую применимость данного метода.

Рассмотренные методы можно комбинировать и дополнять. Для повышения реалистичности синтезируемых ЭКГ-сигналов, позволяющих лучше имитировать характеристики реальных записей, без ручного задания формы волны желательно обучать генеративные модели (GAN, VAE, Diffusion Models) на реальных записях ЭКГ. Однако даже без обучения представленные способы полезны в ряде случаев, когда требуется многократно проверить устойчивость алгоритмов к шумам, редким событиям и смене состояний.

Дополнительно в работе проведена количественная проверка адекватности синтетических сигналов на основе независимых физиологических параметров, не закладываемых напрямую в процесс генерации. В частности, были рассчитаны метрики вариабельности сердечного ритма (HRV) — средняя частота ЧСС, SDNN и RMSSD. Сравнение с реальными данными показало сопоставимость значений: средняя частота ЧСС составила 73 уд./мин для синтетических сигналов и 71 уд./мин для реальных, при этом SDNN и RMSSD оказались даже выше у синтетических данных, что свидетельствует о сохранении естественной динамики ритма и допускает вариативность, необходимую для обучения и тестирования алгоритмов анализа ЭКГ

Список литературы

1. Pan J., Tompkins W. J. A real-time QRS detection algorithm // IEEE Transactions on Biomedical Engineering. 1985. Vol. 32, № 3. P. 230–236. doi:10.1109/TBME.1985.325532
2. Moody G. B., Mark R. G. The impact of the MIT-BIH Arrhythmia Database // IEEE Engineering in Medicine and Biology Magazine. 2001. Vol. 20, № 3. P. 45–50. doi:10.1109/51.932724
3. Behar J., Oster J., Clifford G. D., et al. ECG signal quality during arrhythmia and its application to false alarm reduction // IEEE Transactions on Biomedical Engineering. 2013. Vol. 60, № 6. P. 1660–1666. doi:10.1109/TBME.2013.2240452
4. Laguna P., Mark R. G., Goldberg A., Moody G. B. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG // Proceedings of Computers in Cardiology. 1997. Vol. 24. P. 673–676. doi:10.1109/CIC.1997.648140
5. Sameni R., Clifford G., Jutten C., Shamsollahi M. Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals // EURASIP Journal on Advances in Signal Processing. 2007. Vol. 2007, № 1. P. 94. doi:10.1155/2007/43407
6. Clifford G. D., Azuaje F., McSharry P. E. Advanced methods and tools for ECG data analysis. Norwood, MA: Artech House, 2006.
7. Xue J. Q., Hu Y. H., Tompkins W. J. Neural-network-based adaptive matched filtering for QRS detection // IEEE Transactions on Biomedical Engineering. 1992. Vol. 39, № 4. P. 317–329. doi:10.1109/10.126604
8. Goldberger A. L., Amaral L. A. N., Glass L. et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals // Circulation. 2000. Vol. 101, № 23. P. e215–e220. doi:10.1161/01.CIR.101.23.e215
9. Redmond S. J., Xie Y., Chang D., et al. Electrocardiogram signal quality measures for unsupervised telehealth environments // Physiological Measurement. 2012. Vol. 33, № 9. P. 1517–1533. doi:10.1088/0967-3334/33/9/1517
10. Hamilton P. S., Tompkins W. J. Quantitative investigation of QRS detection rules using the MIT-BIH Arrhythmia Database // IEEE Transactions on Biomedical Engineering. 1986. Vol. 33, № 12. P. 1157–1165. doi:10.1109/TBME.1986.325695
11. Goodfellow I., Pouget-Abadie J., Mirza M., et al. Generative Adversarial Networks // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 2672–2680.
12. Kingma D. P., Welling M. Auto-Encoding Variational Bayes // arXiv:1312.6114, 2013.
13. Isola P., Zhu J. Y., Zhou T., Efros A. A. Image-to-image translation with conditional adversarial networks // Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017. P. 5967–5976. doi:10.1109/CVPR.2017.632
14. Кропачёва А. М., Гурдюк Д. В., Иов И. Л., Першин А. Ю. Генерация временных рядов с простран-

- ственными взаимосвязями // Труды Института системного программирования РАН. 2024. Т. 36, № 1. С. 91–104. doi:10.15514/ispras-2024-36(4)-11
15. Кесиян Г. А., Уртенев М. Х., Шахмеликян Т. А. Анализ методов генерации временных рядов с долговременной корреляционной структурой // Научный журнал КубГАУ. 2011. № 74(10). С. 1–14.
16. Иванча А. Г. Решение задачи генерирования случайных автокоррелированных временных рядов методами непараметрической статистики // Управление экономическими системами: электронный научный журнал. 2011. № 31.

Поступила в редакцию 12.03.2025, окончательный вариант — 21.05.2025.

Калмыков Михаил Александрович, аспирант, СПбГЭТУ «ЛЭТИ», ✉ mica_2011@mail.ru

Шичкина Юлия Александровна, доктор техн. наук, доцент, профессор кафедры вычислительной техники, СПбГЭТУ «ЛЭТИ», strange.y@mail.ru

Computer tools in education, 2025

№ 2: 24–35

<http://cte.eltech.ru>

doi:10.32603/2071-2340-2025-2-24-35

Comparison of Methods for Generating Synthetic Non-Stationary ECG-Like Signals for Testing Time Series Analysis Algorithms

Kalmykov M. A.¹, Postgraduate, ✉ mica_2011@mail.ru
Shichkina Y. A.¹, Doctor sc., Docent, Professor, strange.y@mail.ru,
orcid.org/0000-0001-7140-1686

¹Saint Petersburg Electrotechnical University,
5, building 3, Professora Popova st., 197022, Saint Petersburg, Russia

Abstract

In this paper, various approaches to the generation of synthetic signals simulating a human electrocardiogram (ECG) are considered, with an emphasis on non-stationarity and the presence of various waveforms. Three main types of methods are proposed: 1) rule-based, based on the sum of Gaussians for modeling waves P, Q, R, S, T; 2) stochastic models based on Markov chains, allowing to emulate transitions between different physiological states; 3) neural network generators without strict rules (for example, a recurrent LSTM network with random weights). It is shown how each of the models can be modified to introduce nonstationarity (variations in the duration of cardiac cycles, switching states) and adding local recording artifacts (noisy areas). The proposed methods can be used in testing clustering and time series analysis algorithms when it is necessary to test the methods' resistance to noise, rare events, and state changes.

Keywords: *synthetic ECG, nonstationary signals, signal generation, time series, rule-based methods, Markov chains, neural network generators, LSTM.*

Citation: M. A. Kalmykov and Y. A. Shichkina, "Comparison of Methods for Generating Synthetic Non-Stationary ECG-Like Signals for Testing Time Series Analysis Algorithms,"

Computer tools in education, no. 2, pp. 24–35, 2025 (in Russian); doi:10.32603/2071-2340-2025-2-24-35

References

1. J. Pan and W. J. Tompkins, “A real-time QRS detection algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, Mar. 1985; doi:10.1109/TBME.1985.325532
2. G. B. Moody and R. G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, May-Jun. 2001; doi:10.1109/51.932724
3. J. Behar, J. Oster, G. D. Clifford et al., “ECG signal quality during arrhythmia and its application to false alarm reduction,” *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 6, pp. 1660–1666, Jun. 2013; doi:10.1109/TBME.2013.2240452
4. P. Laguna, R. G. Mark, A. Goldberg, and G. B. Moody, “A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG,” in *Proceedings of Computers in Cardiology*, 1997, pp. 673–676; doi:10.1109/CIC.1997.648140
5. R. Sameni, G. Clifford, C. Jutten, and M. B. Shamsollahi, “Multichannel ECG and noise modeling: Application to maternal and fetal ECG signals,” *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 94, Jan. 2007; doi:10.1155/2007/43407
6. G. D. Clifford, F. Azuaje, and P. E. McSharry, *Advanced Methods and Tools for ECG Data Analysis*. Norwood, MA: Artech House, 2006.
7. J. Q. Xue, Y. H. Hu, and W. J. Tompkins, “Neural-network-based adaptive matched filtering for QRS detection,” *IEEE Transactions on Biomedical Engineering*, vol. 39, no. 4, pp. 317–329, Apr. 1992; doi:10.1109/10.126604
8. A. L. Goldberger, L. A. N. Amaral, L. Glass et al., “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000; doi:10.1161/01.CIR.101.23.e215
9. S. J. Redmond, Y. Xie, D. Chang et al., “Electrocardiogram signal quality measures for unsupervised telehealth environments,” *Physiological Measurement*, vol. 33, no. 9, pp. 1517–1533, Sep. 2012; doi:10.1088/0967-3334/33/9/1517
10. P. S. Hamilton and W. J. Tompkins, “Quantitative investigation of QRS detection rules using the MIT-BIH Arrhythmia Database,” *IEEE Transactions on Biomedical Engineering*, vol. 33, no. 12, pp. 1157–1165, Dec. 1986; doi:10.1109/TBME.1986.325695
11. I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., “Generative adversarial networks,” in *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680, 2014.
12. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” *arXiv:1312.6114*, 2013.
13. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proc. of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017*, pp. 5967–5976; doi:10.1109/CVPR.2017.632
14. A. M. Kropacheva, D. V. Girdyuk, I. L. Iov, and A. Y. Pershin, “Generation of Spatial Time Series Data,” *Proceedings of the Institute for System Programming of the RAS*, vol. 36, no. 4, pp. 143–154, 2024 (in Russian); doi:10.15514/ispras-2024-36(4)-11.
15. G. A. Kesiyan, M. H. Urtenov, and T. A. Shahmelikyan, “Analysis of methods for generating time series with long-term correlation structure,” *Scientific Journal of Kuban State Agrarian University*, no. 74(10), pp. 1–14, 2011 (in Russian).
16. A. G. Ivancha, “Solving the problem of generating random autocorrelated time series using nonparametric statistics methods,” *Management of Economic Systems: Electronic Scientific Journal*, no. 31, 2011 (in Russian).

Received 12-03-2025, the final version — 21-05-2025.

Mikhail Kalmykov, Postgraduate, Saint Petersburg Electrotechnical University “LETI”,
✉ mica_2011@mail.ru

Yulia Shichkina, Doctor of Sciences (Tech.), Docent, Professor, Department of Computer Engineering, Saint Petersburg Electrotechnical University “LETI”, strange.y@mail.ru