

Сравнительная оценка 1D-CNN, CNN-GRU и STFT-ResNet для классификации артефактов движения в носимом фотоплетизмографе

Ч. Ч. Амади

Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
conzira@gmail.com

Ю. О. Боброва

Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
jobobrova@etu.ru

Аннотация. Носимая запястная фотоплетизмография (PPG) обеспечивает непрерывный мониторинг сердечно-сосудистой системы, однако её устойчивость существенно снижается из-за артефактов движения. В данной работе представлена сравнительная оценка глубоких нейросетевых архитектур для классификации двигательных артефактов в запястных PPG-сигналах. Эксперименты проводились на наборе данных PPG-DaLiA, который содержит запястные PPG-сигналы, данные акселерометра и синхронизированные эталонные измерения ЭКГ, полученные в условиях повседневной активности. Поскольку исходные метки «чистый/артефакт» отсутствуют, была предложена консервативная стратегия прокси-разметки, основанная на расхождении частоты сердечных сокращений, определённой по ЭКГ, и статистике интенсивности движения, что позволило сформировать задачу бинарной классификации с окнами «чистый сигнал» и «с артефактами». Были оценены три архитектуры с использованием перекрёстной проверки по схеме leave-one-subject-out (LOSO): одномерная сверточная нейронная сеть (1D-CNN) на исходных PPG-окнах, гибридная модель CNN-GRU на исходных PPG-окнах и модель STFT-ResNet на временно-частотных представлениях. Результаты показывают, что модели, работающие в исходной временной области, превосходят подход на основе преобразованного представления для данной задачи. Модель CNN-GRU продемонстрировала наилучшие результаты: средняя точность 91,22%, сбалансированная точность 90,21%, F1-мера 83,06% и площадь под ROC-кривой (AUC) 0,9726, превзойдя как базовую модель 1D-CNN, так и STFT-ResNet. Наблюдаемые улучшения особенно заметны на более сложных субъектах, что указывает на способность рекуррентного моделирования временной структуры захватывать динамику артефактов, выходящую за пределы чисто сверточного извлечения морфологических признаков. Полученные результаты свидетельствуют о том, что моделирование во временной области более эффективно, чем спектрограммное представление, для посубъектной классификации двигательных артефактов в носимых запястных PPG-сигналах.

Ключевые слова: фотоплетизмография (PPG), классификация двигательных артефактов, носимые сенсорные системы, запястная PPG, глубокое обучение, 1D-CNN, CNN-GRU, STFT-ResNet, посубъектная оценка, PPG-DaLiA

I. ВВЕДЕНИЕ

Фотоплетизмография — это неинвазивный оптический метод измерения изменений объёма крови в микроциркуляторном русле тканей, который стал одной из ключевых сенсорных технологий в системах носимого мониторинга сердечно-сосудистой системы [1], [2]. Запястная PPG особенно привлекательна, поскольку обеспечивает комфортный и длительный мониторинг в естественных условиях. Однако она имеет существенное ограничение: место измерения крайне чувствительно к движениям, смещению датчика, деформации кожи и другим источникам помех, которые искажают форму сигнала и снижают его качество [1], [2]. В практических носимых системах эти помехи не являются незначительным шумом — они представляют собой одно из основных препятствий для надёжного физиологического мониторинга.

Эта проблема особенно важна при амбулаторном мониторинге, когда пользователи выполняют разнообразные повседневные действия, а не остаются неподвижными в лабораторных условиях. По этой причине исследования в области носимой PPG всё больше акцентируют внимание на устойчивости к артефактам реального мира и способности моделей обобщаться на разных пользователях и виды активности [2], [3]. В таких условиях надёжная классификация артефактов может служить важным начальным этапом перед последующей оценкой частоты сердечных сокращений, восстановлением сигнала или отбраковкой данных с учётом их качества. Вместо того чтобы рассматривать все сегменты PPG одинаково, система мониторинга может сначала определить, является ли сигнал достаточно «чистым» или, вероятно, искажённым движением.

Настоящее исследование решает эту задачу путём построения консервативной прокси-задачи классификации «чистый сигнал против сигнала с артефактами» на наборе данных PPG-DaLiA и сравнительного анализа трёх глубоких нейросетевых моделей: 1D-CNN на исходных окнах PPG, CNN-GRU на исходных окнах PPG и STFT-ResNet на временно-частотных представлениях. Цель работы заключается не в полной компенсации артефактов, а в определении наиболее эффективной архитектуры для посубъектной классификации двигательных артефактов в носимой запястной PPG. Такой подход соответствует текущему этапу более широкой исследовательской программы, где

надёжная классификация рассматривается как необходимый предварительный шаг для последующих исследований по компенсации артефактов и анализу сигналов с учётом их качества.

II. МАТЕРИАЛЫ И МЕТОДЫ

A. Набор данных

Эксперименты проводились на наборе данных PPG-DaLiA — публичном датасете, предназначенном для оценки методов определения частоты сердечных сокращений по PPG в условиях повседневной активности с наличием двигательных искажений [3], [4]. Набор данных включает физиологические и двигательные сигналы, зарегистрированные с использованием запястного устройства Empatica E4 и нагрудного устройства RespiBAN. Запястное устройство предоставляет сигнал фотоплетизмографии (blood volume pulse) с частотой дискретизации 64 Гц и трёхосевые данные акселерометра с частотой 32 Гц, тогда как нагрудное устройство обеспечивает ЭКГ и другие физиологические сигналы с частотой 700 Гц [4]. Как описано Reiss et al., PPG-DaLiA представляет собой крупномасштабный набор данных для реальных условий повседневной жизни, содержащий более 36 часов записей от 15 испытуемых и охватывающий широкий спектр видов активности [3]. Сочетание запястного PPG, данных о движении и надёжного ЭКГ-референса делает этот набор данных особенно подходящим для задач анализа двигательных артефактов в PPG-сигналах.

B. Оконная сегментация и преобработка

Чтобы сохранить согласованность с исходной структурой бенчмарка, сигналы были разбиты на окна длиной 8 секунд с шагом 2 секунды. Такой выбор соответствует оригинальному исследованию PPG-DaLiA, в котором использовались 8-секундные окна разметки и посубъектная оценка для анализа способности к обобщению [3]. Таким образом, каждое окно запястной PPG содержало 512 отсчётов, тогда как соответствующее окно акселерометра содержало 256 отсчётов. Для моделей, работающих с исходной временной областью, применялась постоконная нормализация по z-оценке для окон PPG. Для модели, использующей преобразованное представление, каждое окно преобразовывалось в спектрограмму кратковременного преобразования Фурье (STFT) по логарифму амплитуды и затем нормализовалось отдельно для каждой спектрограммы. Использование окон фиксированной длины с перекрытием обеспечило прямую сопоставимость трёх моделей при одинаковом протоколе сегментации и оценки.

B. Стратегия прокси-разметки

Ключевой методологической проблемой является то, что PPG-DaLiA не содержит нативных меток «чистый сигнал/артефакт». Хотя набор данных включает запястную PPG, сигналы движения и ЭКГ-референс для оценки частоты сердечных сокращений, он изначально не был разработан как датасет для классификации качества сигналов [4]. Для построения научно обоснованной задачи бинарной классификации была предложена консервативная стратегия прокси-разметки. Сначала для каждого отсчёта в окне вычислялась величина ускорения акселерометра как:

$$a_{\text{mag}}(n) = \sqrt{a_x^2(n) + a_y^2(n) + a_z^2(n)} \quad (1)$$

где $a_x(n)$, $a_y(n)$ и $a_z(n)$ — три компоненты акселерометра в момент отсчёта n . Для каждого окна i далее была определена характеристика интенсивности движения как стандартное отклонение величины ускорения акселерометра,

$$m_i = \text{std}(a_{\text{mag}}^{(i)}) \quad (2)$$

и нормализована внутри каждого испытуемого для получения субъект-специфического z-значения интенсивности движения.

$$z_i = \frac{m_i - \mu_m}{\sigma_m} \quad (3)$$

где μ_m и σ_m обозначают среднее значение и стандартное отклонение m_i по всем окнам одного и того же испытуемого.

Во-вторых, для каждого PPG-окна была получена оценка частоты сердечных сокращений на основе БПФ. Пусть $X_i(f)$ обозначает спектр Фурье i -го окна PPG. Частота сердечных сокращений, полученная из PPG, оценивалась как доминирующий спектральный пик в физиологически допустимом диапазоне следующим образом:

$$fR^{PPGi} = 60 \cdot \underset{f \in [f_{\min}, f_{\max}]}{\text{argmax}} |X_i(f)| \quad (4)$$

где f_{\min} и f_{\max} задают диапазон поиска, соответствующий физиологически допустимым значениям частоты сердечных сокращений.

Затем расхождение между оценкой, полученной по PPG, и эталонной частотой сердечных сокращений, рассчитанной по ЭКГ, вычислялось как:

$$e_i = |fR^{PPGi} - HR^{ECGi}| \quad (5)$$

где HR^{ECGi} — эталонная частота сердечных сокращений, полученная по ЭКГ для окна i .

Наконец, каждому окну присваивалась прокси-метка качества на основе совместного критерия расхождения частоты сердечных сокращений и интенсивности движения:

$$y_i = \begin{cases} \text{чистый сигнал,} & e_i \leq 5 \text{ и } z_i \leq 0.5 \\ \text{артефакт,} & e_i \geq 10 \text{ и } z_i \geq 1.0 \\ \text{неопределённый,} & \text{иначе} \end{cases} \quad (6)$$

Таким образом, окна с низким расхождением частоты сердечных сокращений и низкой интенсивностью движения были отнесены к классу «чистый сигнал», тогда как окна с высоким расхождением и высокой интенсивностью движения были отнесены к классу «артефакт». Остальные окна были помечены как «неопределённые» и исключены из процесса обучения моделей. Такой консервативный подход был выбран потому, что на начальном этапе исследования задач классификации предпочтительнее использовать меньшие, но более надёжные обучающие выборки, чем большие наборы данных со значительным уровнем шумных меток.

Г. Архитектуры моделей

Были оценены три семейства моделей. Первая модель представляла собой 1D-CNN, работающую непосредственно с исходными окнами PPG. Сверточные слои использовались для извлечения локальной

морфологии сигнала, после чего применялись слои пулинга и полносвязная классификационная «голова». Данная модель использовалась в качестве основного базового варианта, поскольку она является вычислительно эффективной, интерпретируемой на уровне сигнала и хорошо подходит для обучения локальным паттернам формы пульсовой волны.

Вторая модель представляла собой гибридную архитектуру CNN-GRU. На первом этапе сверточный блок извлекал локальные временные признаки, после чего модуль на основе gated recurrent unit (GRU) моделировал более длинные временные зависимости внутри окна. Данная модель была разработана для проверки гипотезы о том, что динамика артефактов лучше улавливается при добавлении рекуррентного последовательностного моделирования поверх сверточного извлечения признаков.

Третья модель представляла собой STFT-ResNet. Каждое исходное окно PPG преобразовывалось в спектрограмму кратковременного преобразования Фурье (STFT) по логарифму амплитуды, которая затем рассматривалась как двумерное изображение и обрабатывалась компактной остаточной сверточной сетью. Данная архитектура использовалась для проверки гипотезы о том, что двигательные артефакты могут быть более эффективно разделены в временно-частотной области, чем в исходной форме временного сигнала. Временно-частотные представления ранее применялись в исследованиях PPG, включая оригинальную структуру Deep PPG, где синхронизированные спектры PPG и акселерометра использовались в качестве входных данных для оценки частоты сердечных сокращений [3].

Д. Параметры обучения

Все модели обучались по одинаковому протоколу leave-one-subject-out с использованием набора данных PPG-DaLiA. Сигналы запястной PPG, дискретизированные с частотой 64 Гц, были разделены на окна длиной 8 секунд с шагом 2 секунды, что соответствовало входным сегментам по 512 отсчётов. Оптимизация моделей выполнялась с использованием оптимизатора Adam с коэффициентом обучения 0.001, при этом для улучшения сходимости и компенсации дисбаланса классов применялись ранняя остановка и взвешивание классов. В каждом фолде валидационная выборка формировалась только из обучающих субъектов, а итоговое сравнение моделей проводилось при фиксированном пороге принятия решения 0.5. Для моделей, работающих с исходными временными сигналами, применялась постоконная нормализация по z-оценке, тогда как для модели на преобразованных данных использовалась нормализация каждой спектрограммы после STFT-преобразования.

Е. Протокол оценки

Все эксперименты проводились с использованием перекрёстной проверки leave-one-subject-out. В каждом фолде один испытуемый использовался в качестве тестовой выборки, а остальные — в качестве обучающего множества. Валидационная подвыборка формировалась исключительно из обучающих субъектов и использовалась для поддержки процедуры ранней остановки. Такая схема оценки является критически важной для носимых физиологических данных, поскольку случайное разбиение на уровне окон может

завышать результаты за счёт смешивания индивидуальных паттернов между обучающей и тестовой выборками. Reiss et al. отдельно подчёркивают необходимость посубъектной оценки при анализе устойчивости и обобщающей способности моделей на PPG-DaLiA [3]. Качество работы моделей оценивалось с использованием метрик точности (accuracy), сбалансированной точности (balanced accuracy), F1-меры и площади под ROC-кривой (AUC). Для базовой модели 1D-CNN дополнительно проводился анализ настройки порога принятия решения на основе валидационной выборки, однако, поскольку полученные улучшения были нестабильными между фолдами, в итоговом сравнении моделей использовался фиксированный порог 0.5.

III. РЕЗУЛЬТАТЫ

А. Анализ разделимости классов меток

Стратегия прокси-разметки обеспечила значимое разделение окон на классы «чистый сигнал», «артефакт» и «неопределённые». Согласно диаграммам рассеяния на рис. 1, окна класса «чистый сигнал» сосредоточены в области с низким значением ошибки и низкой интенсивностью движения, окна с артефактами занимают отдельную область с более высокими значениями ошибки и движения, тогда как окна класса «неопределённые» распределены в промежуточной зоне и включают случаи с противоречивыми характеристиками.



Рис. 1. Диаграмма рассеяния, показывающая классификацию меток

Визуальный анализ репрезентативных примеров подтвердил, что окна класса «чистый сигнал» в целом характеризуются более регулярной пульсирующей морфологией и меньшей вариативностью движения, тогда как окна с артефактами часто демонстрируют искажённую форму сигнала, нестабильное поведение амплитуды и более выраженные колебания, связанные с движением. Эти наблюдения подтверждают практическую обоснованность предложенной стратегии прокси-разметки для начального этапа исследования классификации двигательных артефактов.

Б. Сравнительная эффективность моделей

1D-CNN сформировала сильную базовую модель для работы с исходным сигналом. При оценке по схеме leave-one-subject-out, как показано в табл. 1, она достигла средней точности 0.9112, средней сбалансированной точности 0.8841, средней F1-меры 0.8186 и средней AUC 0.9689. Эти результаты показывают, что локальная морфология формы сигнала уже содержит значительный объём информации, достаточный для различения окон запястной PPG, соответствующих «чистому сигналу» и сигналу, искажённому артефактами.

Модель CNN-GRU показала наилучшие результаты среди всех трёх архитектур. Её средняя точность составила 0.9122, средняя сбалансированная точность — 0.9021, средняя F1-мера — 0.8306, а средняя AUC — 0.9726. По сравнению с базовой моделью 1D-CNN, CNN-GRU обеспечила улучшение сбалансированной точности на 0.0180, F1-меры — на 0.0120 и AUC — на 0.0037. Эти улучшения являются незначительными с точки зрения общей точности, однако более существенными для метрик, учитывающих баланс классов, что особенно важно для задач детектирования артефактов, где распределение классов не является строго сбалансированным.

ТАБЛИЦА I. РЕЗУЛЬТАТЫ РАБОТЫ CNN-МОДЕЛЕЙ

Модель	Accuracy	Balanced Accuracy	F1-мера	AUC
1D-CNN	0.9112	0.8841	0.8186	0.9689
CNN-GRU	0.9122	0.9021	0.8306	0.9726
STFT-ResNet	0.8900	0.8922	0.7945	0.9628

Модель STFT-ResNet в целом не превзошла модели, работающие с исходным временным сигналом. Её средняя точность составила 0.8900, средняя сбалансированная точность — 0.8922, средняя F1-мера — 0.7945, а средняя AUC — 0.9628. По сравнению с 1D-CNN она показала небольшое улучшение сбалансированной точности, однако заметное снижение F1-меры и AUC. По сравнению с CNN-GRU она уступила по всем четырём основным метрикам. Это указывает на то, что использованное временно-частотное представление не сохраняет информацию, релевантную для артефактов, так же эффективно, как прямое моделирование исходной формы сигнала для данной задачи.

В. Наблюдения на уровне отдельных испытуемых

Анализ на уровне отдельных испытуемых показал, что модель CNN-GRU улучшала F1-меру для большинства субъектов и была особенно эффективна на более сложных фолдах, таких как S7, S3, S13 и S9. Этот результат является важным, поскольку он указывает на то, что рекуррентное моделирование временной структуры приносит наибольшую пользу в случаях, когда различие между классами «чистый сигнал» и «артефакт» определяется не только локальной морфологией, но и динамикой сигнала на протяжении всего окна. В отличие от этого, модель STFT-ResNet не демонстрировала столь стабильных преимуществ на уровне отдельных субъектов и обеспечивала менее благоприятный баланс между чувствительностью и специфичностью.

Г. Анализ настройки порога

Также для базовой модели 1D-CNN проводилась настройка порога на основе валидационной выборки. Стратегия с оптимизированным порогом позволила улучшить результаты на некоторых сложных субъектах с высокой долей артефактов, однако эти улучшения не были стабильными во всех фолдах leave-one-subject-out. Оптимальные пороги существенно варьировались между испытуемыми, а в ряде фолдов наблюдалось ухудшение результатов по сравнению с базовым фиксированным порогом. По этой причине настройка порога рассматривалась как вспомогательный экспериментальный анализ, а не использовалась в качестве основной схемы оценки.

IV. ОБСУЖДЕНИЕ

Результаты позволяют сделать три основных вывода. Во-первых, моделирование в исходной временной области эффективно для классификации двигательных артефактов в носимой запястной PPG. Обе модели, работающие с исходным сигналом, показали высокие значения AUC и устойчивую производительность на уровне отдельных испытуемых, что указывает на сохранение в исходной форме волны дискриминативной информации об артефактах без необходимости преобразования в спектрограмму. Это важный практический результат, поскольку он свидетельствует о том, что прямое моделирование временного сигнала не только концептуально проще, но и более эффективно в условиях проведённых экспериментов.

Во-вторых, временная рекуррентность улучшает результаты по сравнению с чисто сверточным базовым подходом. CNN-GRU не обеспечила резкого скачка производительности, однако стабильно улучшала наиболее важные метрики, особенно сбалансированную точность и F1-меру, и была особенно полезна на более сложных испытуемых. Это указывает на то, что двигательные артефакты представляют собой не только локальные искажения формы пульсовой волны, но и включают временную динамику, которую более эффективно улавливает рекуррентный модуль.

В-третьих, протестированный подход на основе временно-частотного представления не оказался наилучшим выбором для данной задачи. Модель STFT-ResNet в целом не превзошла модели, работающие с исходными сигналами, что указывает на то, что преобразование в временно-частотные изображения может приводить к потере или сглаживанию некоторых деталей формы сигнала, которые остаются полезными для различения классов «чистый сигнал» и «артефакт». Это не означает, что методы, основанные на преобразованных представлениях, в целом неэффективны, однако в рамках данной постановки задачи результаты показывают, что представления в виде исходного временного сигнала являются предпочтительными.

Исследование также имеет ряд ограничений. Наиболее важным из них является то, что используемые метки являются эвристическими, а не нативными аннотациями, поскольку PPG-DaLiA изначально является набором данных для оценки частоты сердечных сокращений [4]. Хотя предложенная стратегия прокси-разметки является консервативной и основана на ЭКГ-референсе и признаках движения, определённый уровень

шума в метках остаётся неизбежным. Кроме того, все основные сравнения моделей проводились только на основе запястного PPG-сигнала в качестве входа. Поскольку двигательные артефакты напрямую связаны с физическим движением, в будущих работах следует рассмотреть мультимодальное объединение (fusion) каналов PPG и акселерометра.

V. ЗАКЛЮЧЕНИЕ

Данная работа представила сравнительную оценку моделей 1D-CNN, CNN-GRU и STFT-ResNet для классификации двигательных артефактов в носимой запястной PPG с использованием набора данных PPG-DaLiA. Была предложена консервативная стратегия прокси-разметки для получения обучающих меток «чистый сигнал» и «артефакт» на основе расхождения частоты сердечных сокращений, рассчитанной по ЭКГ, и статистики движения, полученной по данным акселерометра. При оценке по схеме leave-one-subject-out модель CNN-GRU показала наилучшие результаты, превзойдя как базовую модель 1D-CNN, так и STFT-ResNet. Полученные результаты указывают на то, что моделирование во временной области является более эффективным, чем использованное spectrogramное представление, для посубъектной классификации двигательных артефактов в носимой запястной PPG. Также рекуррентное моделирование временной структуры добавляет дополнительную дискриминативную способность по сравнению с чисто сверточным базовым подходом.

Эти результаты создают прочную основу для анализа PPG с фокусом на классификацию двигательных артефактов. Они также формируют предпосылки для более широких исследований, связанных с анализом контактных или обусловленных сцеплением артефактов, а также для последующей разработки методов компенсации сигналов с учётом наличия артефактов.

СПИСОК ЛИТЕРАТУРЫ

- [1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiol. Meas.*, vol. 28, no. 3, pp. R1–R39, 2007, doi: 10.1088/0967-3334/28/3/R01.
- [2] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," *Int. J. Biosens. Bioelectron.*, vol. 4, no. 4, pp. 195–202, 2018, doi: 10.15406/ijbsbe.2018.04.00125.
- [3] A. Reiss, I. Indlekofer, P. Schmidt, and K. Van Laerhoven, "Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks," *Sensors*, vol. 19, no. 14, art. 3079, 2019, doi: 10.3390/s19143079.
- [4] A. Reiss, I. Indlekofer, and P. Schmidt, PPG-DaLiA [Dataset]. UCI Machine Learning Repository, 2019, doi: 10.24432/C53890.
- [5] M. Elgendi, "On the analysis of fingertip photoplethysmogram signals," *Curr. Cardiol. Rev.*, vol. 8, no. 1, pp. 14–25, 2012, doi: 10.2174/157340312801215782.