

Physio-JERA для ОРИТ-сигналов: временная мультимодальная объяснимость и метрика *Temporal Range Coherence* в задаче валидации тревог желудочковой тахикардии

Е. М. Кузнецов¹, А. Н. Аверкин¹, Ю. В. Трофимов^{1,2},
М. А. Лопатин¹, Е. С. Кондрашова¹

¹Государственный университет «Дубна»

²Объединенный институт ядерных исследований

E-mail: Kot454556@yandex.ru

Аннотация. Работа посвящена задаче мультимодальной объяснимости трёхканальных физиологических сигналов ОРИТ (ЭКГ, ФПГ, ИАД) при различении истинных и ложных тревог желудочковой тахикардии на базе VTaC. В качестве базовой меры согласованности объяснений используется метрика *Coherence*, основанная на временных атрибуциях *Integrated Gradients* и попарной взаимной информации между модальностями. Предлагается архитектурное и методологическое развитие этого подхода на основе принципов *Joint-Embedding Predictive Architecture*: для физиологических временных рядов вводится модель *Physio-JERA* с модально-специфическими преобразователями для ЭКГ, ФПГ и ИАД, межмодальным предсказанием скрытых временных сегментов и последующим дообучением в задаче классификации тревог. Вводится метрика *Temporal Range Coherence*, обобщающая *Coherence* на диапазон допустимых временных сдвигов между модальностями. Это позволяет учитывать физиологически правдоподобные задержки между электрической активностью сердца и гемодинамическим откликом и тем самым делает анализ объяснений более интерпретируемым с клинической точки зрения.

Ключевые слова: мультимодальная объяснимость, физиологические временные ряды, валидация тревог желудочковой тахикардии, истинные и ложные тревоги ОРИТ, VTaC, *Integrated Gradients*, *self-supervised learning*, *Joint-Embedding Predictive Architecture*, кросс-модальное предсказание, темпоральная согласованность, *Temporal Range Coherence*

I. ВВЕДЕНИЕ И ПОСТАНОВКА ПРОБЛЕМЫ

Многие методы мультимодальной объяснимости исходят из строгой синхронности между модальностями. Однако в условиях ОРИТ такое допущение не всегда физиологически корректно: электрическая активность сердца, регистрируемая по ЭКГ, и гемодинамический отклик, наблюдаемый по ФПГ и ИАД, разделены естественной временной задержкой [1]. Поэтому пообъектное сравнение модальностей может занижать их реальную согласованность. На рис. 1 показан типичный пример такого временного рассогласования.

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации (тема № 124112200072-2)

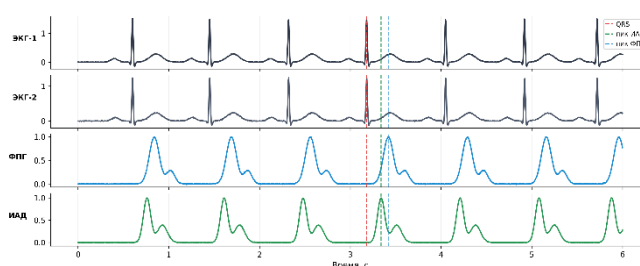


Рис. 1. Физиологическая задержка гемодинамического отклика относительно электрической активности сердца: пик QRS-комплекса на ЭКГ предшествует соответствующим изменениям на ФПГ и ИАД.

Ложные тревоги желудочковой тахикардии остаются одной из наиболее серьёзных проблем клинического мониторинга в отделениях реанимации и интенсивной терапии [2, 3]. Значительная часть таких тревог обусловлена не истинной аритмией, а двигательными артефактами, нарушением контакта электродов или шумами регистрации [1]. Это создаёт избыточную нагрузку на медицинский персонал, снижает доверие к мониторинговым системам и способствует развитию эффекта усталости от тревог [2, 3].

В клинической практике валидация тревоги носит мультимодальный характер. Врач интерпретирует ЭКГ не изолированно, а сопоставляет её с гемодинамическими каналами, прежде всего с фотоплетизмограммой и кривой инвазивного артериального давления. Если ЭКГ демонстрирует признаки тахикардии, а пульсовая и прессиорная динамика остаются стабильными, такое рассогласование скорее указывает на ложную тревогу [1]. Поэтому для интеллектуальных систем мониторинга недостаточно одной лишь высокой точности классификации; модель должна обеспечивать кросс-модальную проверку собственного решения.

Работы по фильтрации ложных тревог показали, что при валидации тревог важную роль играет сопоставление ЭКГ с гемодинамическими сигналами [1]. Позднее эта идея была перенесена в задачи мультимодальной объяснимости, где для оценки межмодальной согласованности временных объяснений, полученных методом *Integrated Gradients*, была

предложена метрика *Coherence* [4, 5]. Было показано, что она может служить дополнительным индикатором надёжности предсказания. Однако исходная постановка *Coherence* основывалась на синхронном сравнении атрибуций и потому не учитывала физиологически правдоподобные временные задержки между модальностями. Иными словами, объяснения оценивались синхронно, хотя сами физиологические процессы разворачиваются с естественной временной задержкой.

Это ограничение устраняется в работе двумя шагами. Во-первых, вводится архитектура *Physio-JEPA* для самоконтролируемого обучения совместных латентных представлений по гетерогенным ОРИТ-сигналам. Во-вторых, предлагается метрика *Temporal Range Coherence*, обобщающая исходную *Coherence* на диапазон допустимых временных сдвигов между модальностями [6]. Тем самым становится возможным перейти от жёсткой синхронной оценки к более физиологически правдоподобной временной согласованности объяснений. Общая схема предлагаемого мультимодального пайплайна приведена на рис. 2.

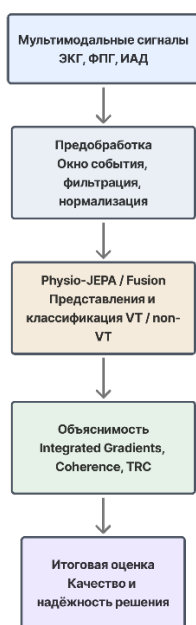


Рис. 2. Общая схема мультимодального пайплайна: от физиологических сигналов к классификации, объяснимости и итоговой оценке надёжности решения.

II. АРХИТЕКТУРА PHYSIO-JEPA

Обучение с учителем в задаче валидации тревог желудочковой тахикардии ограничено объёмом и качеством разметки. Поэтому сначала используется самоконтролируемое предобучение на размеченных мультимодальных записях, а затем выполняется дообучение на целевой задаче классификации. В качестве основы выбрана архитектурная схема семейства *JEPA*, в котором модель предсказывает скрытые участки латентного представления по наблюдаемому контексту, не восстанавливая входной сигнал в исходном пространстве [6]. Дополнительной мотивацией служит то, что для биомедицинских данных мультимодальное объединение источников рассматривается как естественный способ построения более устойчивых и информативных представлений [7].

Упрощённая схема классификационного модуля приведена на рис. 3.

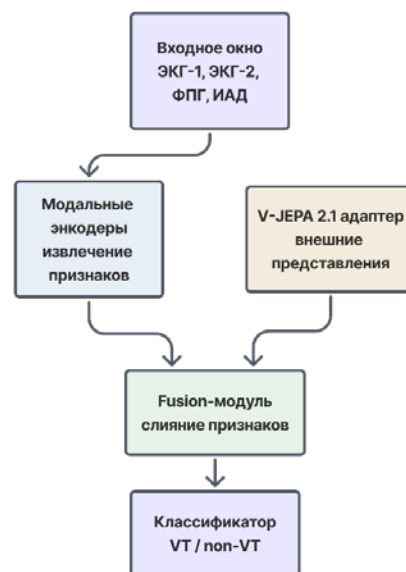


Рис. 3. Схема классификационного модуля с модальными энкодерами, fusion-слиянием и опциональной интеграцией внешних представлений

Пусть после предварительной синхронизации и оконного выделения эпизод задаётся как

$$X = (x^{ECG}, x^{PPG}, x^{ABP}), \quad (1)$$

где x^{ECG} содержит два канала ЭКГ, а x^{PPG} и x^{ABP} соответствуют фотоплетизмограмме и инвазивному артериальному давлению. Для каждой модальности используется собственный модуль преобразования сигнала в последовательность латентных представлений. Далее эти представления объединяются в общее пространство, в котором скрытые фрагменты одной модальности предсказываются по наблюдаемому внутримодальному и межмодальному контексту. Например, скрытый участок ФПГ может восстанавливаться по двум каналам ЭКГ, ИАД и соседним видимым фрагментам самой ФПГ. Такая постановка снижает вероятность того, что модель будет опираться лишь на локальные тривиальные корреляции, и побуждает её использовать более содержательные межмодальные зависимости, в частности физиологические связи между электрической активностью сердца и гемодинамическим откликом. После предварительного самоконтролируемого обучения полученные представления используются на этапе дообучения для бинарной классификации тревог. Таким образом, *Physio-JEPA* служит механизмом формирования латентного пространства, в котором межмодальные зависимости учитываются ещё до этапа целевого обучения.

III. ОБЪЯСНИМОСТЬ И МЕТРИКА TEMPORAL RANGE COHERENCE

Для интерпретации решений модели используются временные атрибуции, получаемые методом *Integrated Gradients* [8]. В отличие от исходной *Coherence*, основанной на синхронном сравнении атрибуций, в данной работе вводится метрика *Temporal Range Coherence (TRC)*, учитывающая допустимые временные сдвиги между модальностями.

$$\tau \in [\tau_{\min}, \tau_{\max}], \quad (2)$$

где границы диапазона выбираются в соответствии с физиологически правдоподобными задержками между электрической активностью сердца и периферическим гемодинамическим откликом [9].

Для каждой пары модальностей типа «ЭКГ–ФПГ» и «ЭКГ–ИАД» согласованность атрибуций оценивается при временном сдвиге τ . В качестве меры согласованности используется взаимная информация между временными картами атрибуций, а сама метрика определяется как максимум по допустимому диапазону временных сдвигов:

$$C_{\tau}(X) = \frac{1}{2} [MI(a^{ECG}(X), S_{\tau}a^{PPG}(X)) + MI(a^{ECG}(X), S_{\tau}a^{ABP}(X))] \quad (3)$$

$$TRC(X) = \max_{\tau \in [\tau_{\min}, \tau_{\max}]} C_{\tau}(X).$$

Здесь $a^{ECG}(X)$, $a^{PPG}(X)$ и $a^{ABP}(X)$ — временные карты атрибуций, $MI(\cdot, \cdot)$ — оператор временного сдвига атрибуций на величину τ в пределах допустимого окна перекрытия [10].

Если решение модели действительно опирается на физиологически связанные события, максимум согласованности должен достигаться при ненулевом, но правдоподобном временном сдвиге, соответствующем задержке между ЭКГ и гемодинамикой. В этом случае TRC принимает высокие значения. Если же решение в основном определяется синхронными артефактами, шумом или аппаратной наводкой, не сопровождаемыми согласованным гемодинамическим ответом, временная согласованность снижается, и значение TRC уменьшается.

IV. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА И ВАЛИДАЦИЯ КОНЦЕПТА

Экспериментальная часть работы направлена на проверку идеи мультимодальной объяснимости с учётом временных сдвигов. Ключевой вопрос состоит в следующем: позволяет ли учёт физиологически допустимых временных сдвигов лучше различать истинные и ложные тревоги по сравнению с синхронной оценкой межмодальной согласованности.

На первом этапе был выполнен анализ распределения временных сдвигов, при которых межмодальная согласованность атрибуций достигает максимума. Для каждого эпизода определялось значение τ^* , при котором метрика TRC принимала наибольшее значение в заданном физиологическом диапазоне. Полученные распределения для истинных и ложных тревог представлены на рис. 4. Характерные значения τ^* не сводятся к строго синхронному случаю $\tau = 0$, а сосредоточены в области ненулевых временных сдвигов. Это показывает, что межмодальная согласованность объяснений в данной задаче не сводится к строго синхронной постановке.

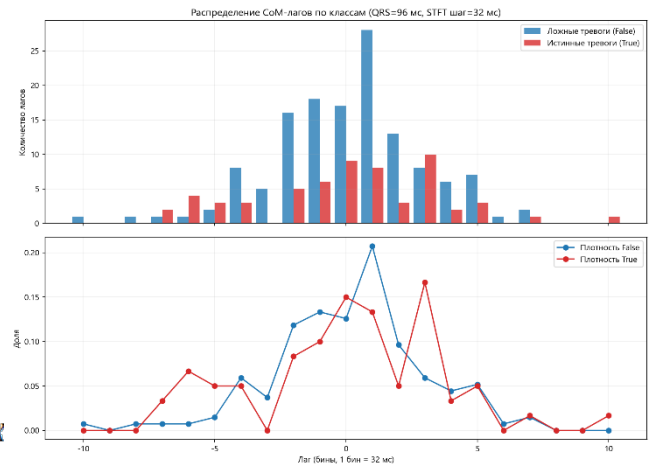


Рис. 4. Распределение межмодальных лагов, соответствующих максимуму согласованности атрибуций, для истинных и ложных тревог

Следующим этапом выполняется сравнение исходной метрики *Coherence* и предложенной метрики TRC с точки зрения их разделяющей способности на выборке истинных и ложных тревог. Предварительные результаты указывают, что для истинных эпизодов тахикардии физиологическая связь между модальностями сохраняется, вследствие чего учёт допустимой временной задержки закономерно максимизирует согласованность атрибуций. В то же время на ложных тревогах, где доминируют синхронные двигательные артефакты, учёт допустимого временного сдвига не сопровождается существенным ростом TRC.

V. ЗАКЛЮЧЕНИЕ

В работе предложен подход к анализу ОРИТ-сигналов, объединяющий самоконтролируемое предиктивное обучение мультимодальных физиологических представлений и интерпретацию решений модели с учётом временных сдвигов. В отличие от синхронной постановки межмодальной согласованности, метрика *Temporal Range Coherence* учитывает физиологически правдоподобные временные задержки между электрической активностью сердца и гемодинамическим откликом, что делает анализ объяснений более интерпретируемым с клинической точки зрения.

Полученные результаты показывают, что учёт временных сдвигов позволяет, более физиологически обоснованно, анализировать межмодальную согласованность объяснений и в ряде случаев выявлять временную структуру связей, не сводимую к строго синхронному сопоставлению, тогда как архитектура *Physio-JEPA* может рассматриваться как средство формирования латентных представлений, учитывающих связи между модальностями.

Дальнейшие исследования связаны с масштабированием самоконтролируемого предварительного обучения на более крупных массивах физиологических данных, включая базы семейства *MIMIC*, а также с изучением возможности включения метрики TRC в процедуру дообучения. Это позволит перейти от последующей оценки согласованности к построению моделей, изначально чувствительных к физиологически правдоподобной межмодальной динамике.

СПИСОК ЛИТЕРАТУРЫ

- [1] Zong W., Moody G. B., Mark R. G. Reduction of false arterial blood pressure alarms using signal quality assessment and relationships between the electrocardiogram and arterial blood pressure // *Medical & Biological Engineering & Computing*. 2004. Vol. 42, no. 5. P. 698–706.
- [2] Drew B. J., Harris P., Zègre-Hemsey J. K., Mammone T., Schindler D., Salas-Boni R., Bai Y., Tinoco A., Ding Q., Hu X. Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients // *PLoS ONE*. 2014. Vol. 9, no. 10. Art. e110274.
- [3] Sendelbach S., Funk M. Alarm Fatigue: A Patient Safety Concern // *AACN Advanced Critical Care*. 2013. Vol. 24, no. 4. P. 378–386.
- [4] Трофимов Ю. В., Аверкин А. Н., Кузнецов Е. М., Еремеев А. П., Нечаевский А. В. Мультимодальная объяснимость для ОРИТ-сигналов (VTaC): метрические и асимптотические результаты // *Вестник ТвГУ. Серия: Прикладная математика*. 2025. № 4. С. 43–80
- [5] Sundararajan M., Taly A., Yan Q. Axiomatic Attribution for Deep Networks // *Proceedings of the 34th International Conference on Machine Learning*. 2017. Vol. 70. P. 3319–3328.
- [6] Assran M., Duval Q., Misra I., Bojanowski P., Vincent P., Rabat M., LeCun Y., Ballas N. Self-Supervised Learning From Images With a Joint-Embedding Predictive Architecture // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023. P. 15619–15629.
- [7] Stahlschmidt S. R., Pahlevan N., Atyabi A., Zoroufian P., Tavallali P., Blaber A. P., Doyle T. E., Reisman D. Multimodal deep learning for biomedical data fusion: a review // *Frontiers in Digital Health*. 2022. Vol. 4. Art. 821871.
- [8] Sundararajan M., Taly A., Yan Q. Axiomatic Attribution for Deep Networks // *Proceedings of the 34th International Conference on Machine Learning*. 2017. Vol. 70. P. 3319–3328.
- [9] Ding X., Zhang Y.-T. Pulse transit time technique for cuffless unobtrusive blood pressure measurement: from theory to algorithm // *Biomedical Engineering Letters*. 2019. Vol. 9, no. 1. P. 37–52.
- [10] Kraskov A., Stögbauer H., Grassberger P. Estimating mutual information // *Physical Review E*. 2004. Vol. 69, no. 6. Art. 066