

## ЛИНГВИСТИЧЕСКИЙ ПОДХОД К АНАЛИЗУ СИГНАЛОВ НА ПРИМЕРЕ ЭКГ

**Файнзильберг Леонид Соломонович,**  
доктор технических наук, профессор,  
профессор кафедры биомедицинской  
кибернетики, Национальный технический  
университет Украины «Киевский  
политехнический институт  
имени Игоря Сикорского», Украина

**Дыкач Юлия Романовна**  
студент VI курса  
Национальный технический  
университет Украины «Киевский  
политехнический институт  
имени Игоря Сикорского», Украина

Электрокардиографию уже более ста лет широко используют в кардиологической практике для диагностики заболеваний сердечно-сосудистой системы. Однако известно, что традиционный подход к анализу и интерпретации ЭКГ не всегда обеспечивает требуемую достоверность принятия диагностических решений. Поэтому специалисты активно исследуют новые подходы к компьютерной обработке ЭКГ.

В работе [1] нами предложен эффективный метод оценки динамики изменений формы циклов ЭКГ, основанный на привлечении лингвистического подхода к обработке циклических сигналов. Этот подход предполагает переход от наблюдаемой ЭКГ к последовательности символов (слову), которое однозначно кодирует ЭКГ. Алгоритм такой кодировки сводится к следующему:

Оценивается динамика двух показателей в процессе регистрации ЭКГ: традиционного показателя продолжительность RR-интервала и оригинального показателя  $\beta_T$ , характеризующего симметрию участка реполяризации на фазовой плоскости. Для этого вводятся индикаторные переменные

$$V_i^{(RR)} = \begin{cases} +1, & \text{если } RR_i - RR_{i-1} > 0, \\ -1, & \text{если } RR_{i-1} - R_i > 0. \end{cases}$$

$$V_i^{(\beta)} = \begin{cases} +1, & \text{если } \beta_{T,i} - \beta_{T,ii-1} > 0, \\ -1, & \text{если } \beta_{T,ii-1} - \beta_{T,i} > 0, \end{cases}$$

где  $i = 2, \dots, N$ .

Последовательности  $V_i^{(RR)}$  и  $V_i^{(\beta)}$  позволяют закодировать каждый цикл ЭКГ одним из символов алфавита  $A = \{a, b, c, d\}$  следующим образом (табл 1).

Таблица 1.  
Принцип кодирования цикла ЭКГ

Значения индикаторной переменной $V_i^{(RR)}$	+1	+1	-1	-1
Значения индикаторной переменной $V_i^{(\beta_T)}$	+1	-1	+1	-1
Символ	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>

Далее для анализа в качестве параметров предлагается использовать не показатели, а кодограммы, а в качестве метрики предлагается использовать расстояние Левенштейна, представляющее собой минимальное количество операций редактирования (вставки, удаления и замены), и обеспечивающее переход от одного слова ко второму.

Пусть в результате экспериментов зарегистрировано  $Q_{CAD}$  электрокардиограмм больных ишемической болезнью сердца ИБС (**CAD**), которые в соответствии с таблицей 1 закодированы словами  $S_q^{(CAD)}$ ,  $q = 1, \dots, Q_{CAD}$ . Определим расстояния Левенштейна  $L_{\mu\nu}(S_\mu^{(CAD)}, S_\nu^{(CAD)})$  между каждой парой  $S_\mu^{(CAD)}, S_\nu^{(CAD)}$ ,  $\mu = 1, \dots, Q_{CAD}$ ,  $\nu = 1, \dots, M_1$  указанных слов и сформируем квадратную  $Q_{CAD} \times Q_{CAD}$  матрицу расстояний  $L_{\mu\nu}(S_\mu^{(CAD)}, S_\nu^{(CAD)})$ ,  $\mu = 1, \dots, Q_{CAD}$ ,  $\nu = 1, \dots, Q_{CAD}$ .

Тогда эталонное слово  $S_0^{(CAD)}$  больного определит строка указанной матрицы, сумма элементов которой минимальна, т.е.

$$S_0^{(CAD)} = \arg \max_{1 \leq \nu \leq Q_{CAD}} \sum_{\mu=1}^{Q_{CAD}} L_{\mu\nu}.$$

Эталонное слово группы здоровых  $S_0^{(Healthy)}$  определяют аналогичным образом. Сформируем следующее решающее правило для классификации:

$$\text{ИБС, если } L(S_t, S_0^{(CAD)}) \leq L(S_t, S_0^{(Healthy)}),$$

$$\text{Здоров, если } L(S_t, S_0^{(CAD)}) > L(S_t, S_0^{(Healthy)}).$$

Клинические исследования показали, что данное правило обеспечивает чувствительность 72% и специфичность 79% даже на тех электрокардиограммах, на которых отсутствуют общепринятые признаки ишемии миокарда.

Для повышения чувствительности и специфичности предложено усовершенствованное решающее правило [2], основанное на анализе вероятности появления символов в кодовых словах. С этой целью нами предложено использовать частоты появления в кодовых словах трех типов паттернов [3]:

$$\pi_1 = x, \quad x \in \{a, b, c, d\} \text{ – односимвольный паттерн;}$$

$\pi_2 = xy$ ,  $x, y \in \{a, b, c, d\}$  – двухсимвольный паттерн;

$\pi_3 = xyz$ ,  $x, y, z \in \{a, b, c, d\}$  – трехсимвольный паттерн.

Это позволило сформулировать усовершенствованное решающее правило:

ИБС, если  $L(S_t, S_0^{CAD}) \leq L(S_t, S_0^{Healthy})$  и  $|p_i(x) - p_{mean}^{CAD}(x)| \leq |p_i(x) - p_{mean}^{healthy}(x)|$ ,

Здоров, если  $L(S_t, S_0^{CAD}) \geq L(S_t, S_0^{Healthy})$  и  $|p_i(x) - p_{mean}^{CAD}(x)| > |p_i(x) - p_{mean}^{healthy}(x)|$ ,

где  $x$  – характерный паттерн,  $p$  – частота вхождения,  $L$  – расстояние Левенштейна между кодограммой и эталонами.

Статистическая обработка клинических данных показала, что использование комплексного анализа, основанного на метрике Левенштейна и частотных оценках, для односимвольного паттерна обеспечивает чувствительность 77,2% и специфичность 86,2%,

**Выводы:** Лингвистический подход, основанный на сравнении расстояний Левенштейна между кодовым словом анализируемой ЭКГ и заранее построенными эталонами верифицированных больных ишемической болезнью сердца и здоровых волонтеров, обеспечивает принятие диагностических решений с чувствительностью 72% и специфичностью 79% даже на тех ЭКГ, когда традиционный ЭКГ анализ оказывается неинформативным.

Усовершенствованное решающее правило, основанное на дополнительном анализе вхождения характерного символьного паттерна в кодовое слово, обеспечивает повышение чувствительности и специфичности диагностических решений.

### Список литературы

1. Fainzilberg L., Dykach J. Linguistic approach for estimation of electrocardiograms's subtle changes based on the Levenshtein distance // Scientific journal «Cybernetics and computer engineering». 2019. № 2. P.3-26. DOI: [https:// 10.15407/kvt196.02.003](https://10.15407/kvt196.02.003).

2. Сенкевич Ю.И. Лингвистический анализ физиологических сигналов // Цифровая обработка сигналов. 2008. № 2. С.54-57.

3. Fainzilberg L., Dykach J. Development of a linguistic approach to the problem of the computer electrocardiogram's classifications // Scientific journal «Control Systems and Computers». 2021. № 2 (в печати).