

Обнаружение маммограмм с патологиями с применением интерфейса «МОЗГ-КОМПЬЮТЕР»

Валентина Сулимова

vsulimova@yandex.ru

ТулГУ, Тула, Россия

Ольга Красоткина

MPI, Summit, USA

Сергей Бухонов

МФТИ, Москва, Россия

Вадим Моттль

ВЦ РАН, Москва, Россия

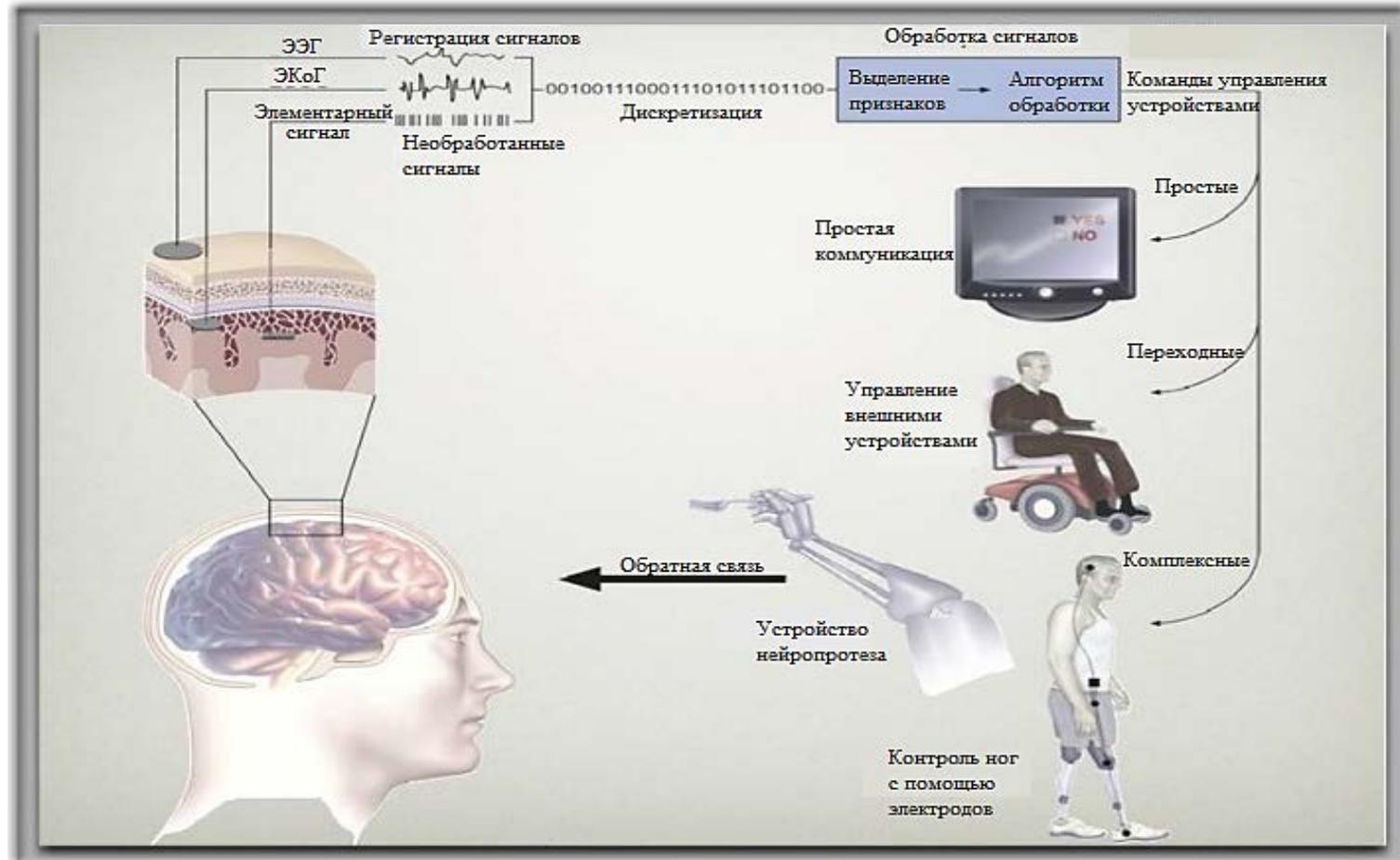
Дэвид Уиндрид

Middlesex University,

London, UK

ИОИ-2018, Гаэта, Италия, октябрь 2018

Применение интерфейса «мозг-компьютер» (ИМК) в реабилитационной медицине

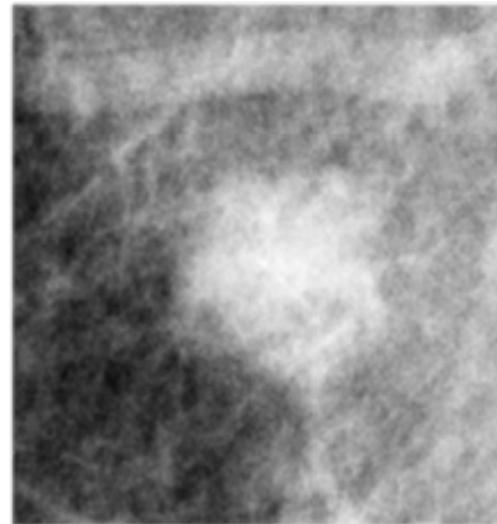


Построение ИМК основано на регистрации электрической активности различных участков головного мозга в виде электроэнцефалограмм и их последующем анализе

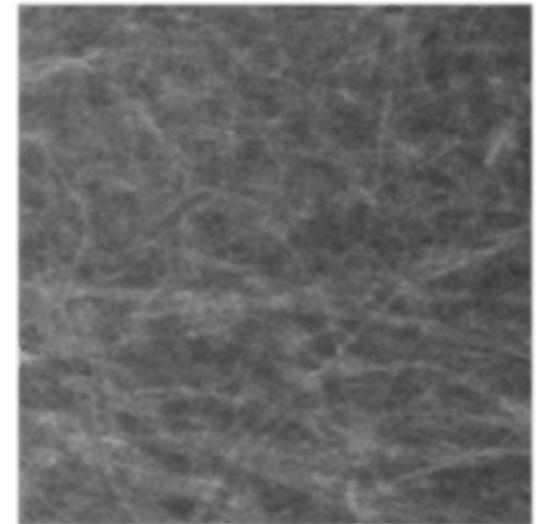
Применение ИМК для высокопроизводительного обнаружения маммограмм с патологиями



*Регистрация ЭЭГ
в процессе просмотра
быстро сменяющихся
изображений маммограмм*



*пример целевого объекта
(маммограмма с
патологией)*

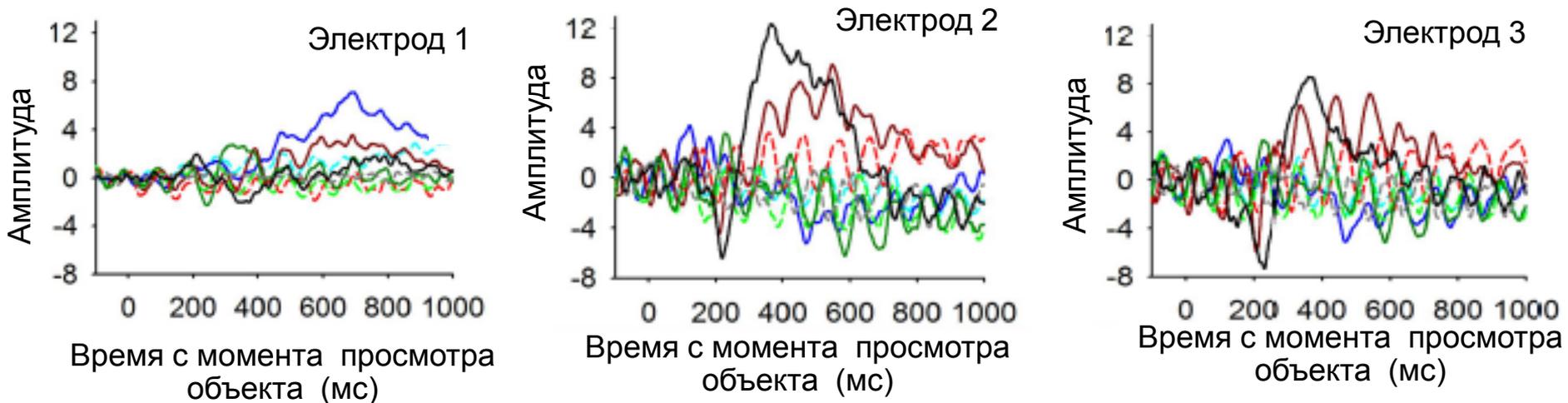


*пример нецелевого объекта
(маммограмма без
патологий)*

C.Hope, A. Sterr, P.E. Langovan, N.Geades, D.Windridge, K.Young, K.Wells.

Throughput Screening for Mammography using a Human-Computer Interface with Rapid Serial Visual Presentation (RSVP) / Proc. SPIE 8673, Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment, 867303 (March 28, 2013); doi:10.1117/12.2007557

Примеры усредненных* ЭЭГ сигналов при просмотре маммограмм несколькими испытуемыми



Сплошные линии – ЭЭГ нескольких испытуемых при просмотре маммограмм **с патологиями**

Пунктирные линии – ЭЭГ нескольких испытуемых при просмотре маммограмм **без патологий**

* Приведенные сигналы – результат усреднения сигналов, зарегистрированных при просмотре нескольких маммограмм

Схема эксперимента

Полная ЭЭГ:

Эксперт просматривает серию изображений маммограмм (по 100 мс на изображение)

Активность его мозга регистрируется на ЭЭГ (66 электродов)

Все сигналы подвергаются низкочастотной фильтрации с частотой среза 40Гц

Задача обнаружения маммограмм с патологиями как задача двухклассового распознавания

Объекты распознавания:

фрагменты полной ЭЭГ (по каждому электроду отдельно) с характеристиками :

- длина 1100 отсчетов (соответствует просмотру 11 маммограмм)

- **для нецелевых объектов:**

все 11 маммограмм без патологий

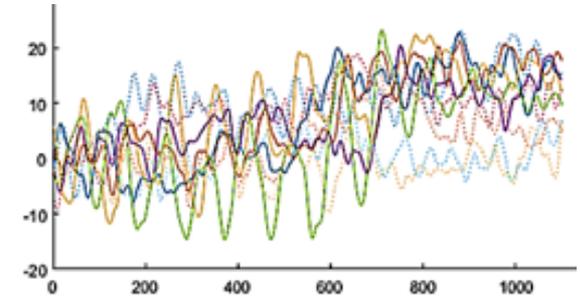
- **для целевых объектов:**

1 из 11 маммограмм с патологией, остальные без патологий
(позиция маммограммы с патологией - случайная с 4 по 7)

Предварительная обработка фрагментов ЭЭГ

До предварительной обработки

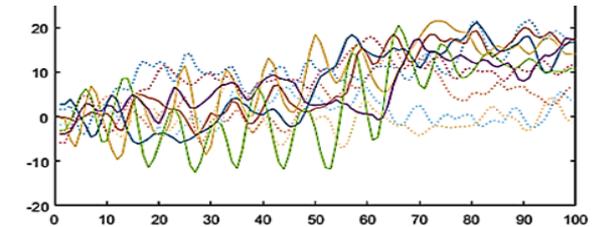
$$\mathbf{x} = (x_i \in R, i = 1, \dots, m), \quad m = 1100$$



Прореживание в $\text{step}=11$ раз

$$\mathbf{x}' = (x'_i \in R, i = 1, \dots, m') \quad x'_i = x_{(i-1) \cdot \text{step} + 1},$$

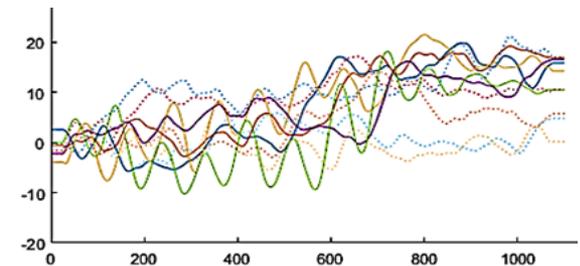
$$m' = m / \text{step} = 1100 / 11 = 100$$



Сглаживание скользящим окном ширины $w=11$

$$x''_i = \frac{1}{w} \sum_{k=i-\lfloor w/2 \rfloor}^{i+\lfloor w/2 \rfloor} x'_k, \quad i = \lfloor w/2 \rfloor + 1, \dots, m' - \lfloor w/2 \rfloor$$

$$x''_i = x'_{\lfloor w/2 \rfloor + 1}, \quad i < \lfloor w/2 \rfloor + 1 \quad x''_i = x'_{m' - \lfloor w/2 \rfloor}, \quad i > m' - \lfloor w/2 \rfloor$$



Метод опорных векторов (SVM) для обнаружения маммограмм с патологией по фрагментам ЭЭГ

Обучающее множество фрагментов ЭЭГ: $\{\mathbf{x}_j, y_j\}$, $\mathbf{x}_j = \mathbf{x}_j(\omega)$, $j = 1, \dots, N$

$y_j = +1$ - есть маммограмма с патологией (целевой класс)

$y_j = -1$ - нет маммограмм с патологией (не целевой класс)

Решающее правило в виде линейной разделяющей гиперплоскости:

$$d(\mathbf{x}; \mathbf{a}, b) = \mathbf{a}^T \mathbf{x} + b \quad \begin{cases} \geq 0 \Rightarrow \hat{y}(\mathbf{x}) = +1, \\ < 0 \Rightarrow \hat{y}(\mathbf{x}) = -1, \end{cases} \quad \begin{array}{l} \mathbf{a} \in R^m \text{ - направляющий вектор} \\ b \text{ - смещение вдоль направляющего вектора} \end{array}$$

Задача нахождения оптимальной разделяющей гиперплоскости:

$$\begin{cases} \sum_{i=1}^m a_i^2 + C \sum_{j=1}^N \delta_j \rightarrow \min(\mathbf{a}, b, \delta), \\ y_j (\sum_{i=1}^m a_i x_{ij} - b) \geq 1 - \delta_j, \quad j = 1, \dots, N, \\ \delta_j \geq 0, \quad j = 1, \dots, N, \end{cases}$$

Параметры оптимальной гиперплоскости в терминах множителей Лагранжа $\lambda_j, j = 1, \dots, N$:

$$\mathbf{a} = \sum_{j=1}^N \lambda_j y_j \mathbf{x}_j \quad b = \frac{1}{2} \left[\min_{j: y_j=1} \sum_{k=1}^N \lambda_k y_k \mathbf{x}_j^T \mathbf{x}_k - \max_{j: y_j=-1} \sum_{k=1}^N \lambda_k y_k \mathbf{x}_j^T \mathbf{x}_k \right]$$

*SVM для обнаружения маммограмм с патологией с требованием гладкости решающего правила**

Обучающее множество фрагментов ЭЭГ: $\{\mathbf{x}_j, y_j\}$, $\mathbf{x}_j = \mathbf{x}_j(\omega)$, $j = 1, \dots, N$

$y_j = +1$ - есть маммограмма с патологией (целевой класс)

$y_j = -1$ - нет маммограмм с патологией (не целевой класс)

Решающее правило в виде линейной разделяющей гиперплоскости:

$$d(\mathbf{x}; \mathbf{a}, b) = \mathbf{a}^T \mathbf{x} + b \quad \begin{cases} \geq 0 \Rightarrow \hat{y}(\mathbf{x}) = +1, \\ < 0 \Rightarrow \hat{y}(\mathbf{x}) = -1, \end{cases} \quad \begin{array}{l} \mathbf{a} \in R^m \text{ - направляющий вектор} \\ b \text{ - смещение вдоль направляющего вектора} \end{array}$$

Задача нахождения оптимальной разделяющей гиперплоскости:

$$\begin{cases} \sum_{i=1}^m a_i^2 + \gamma \sum_{i=2}^m (a_i - a_{i-1})^2 + C \sum_{j=1}^N \delta_j \rightarrow \min(\mathbf{a}, b, \delta), \\ y_j (\sum_{i=1}^m a_i x_{ij} - b) \geq 1 - \delta_j, \quad j = 1, \dots, N, \\ \delta_j \geq 0, \quad j = 1, \dots, N, \end{cases} \quad \begin{array}{l} \gamma \text{ - параметр, определяющий} \\ \text{степень гладкости} \\ \text{решающего правила} \end{array}$$

* Данный способ регуляризации впервые предложен в работе Mottl V.V., Dvoenko S.D., Seredin O.S., Krasotkina O.V. Pattern recognition learning taking into account the criterion of smoothness of the decision rule. // Control and Information: proc. of the chair of automation and remote control of TSU. (2000)

Объединение информации с нескольких электродов

Сигналы с k электродов:

$$\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_m^{(i)}] \in R^m, \quad i = 1, \dots, k$$

Объединенный сигнал:

$$\mathbf{x} = [x_1^{(1)}, \dots, x_m^{(1)}, \dots, x_1^{(k)}, \dots, x_m^{(k)}] = [x_1, \dots, x_{m \cdot k}] \in R^{m \cdot k}$$

При $m = 100$, $k = 66$: $\mathbf{x} \in R^{6600}$

SVM для обнаружения маммограмм с патологией с отбором наиболее информативных отсчетов ЭЭГ*

Задача нахождения оптимальной разделяющей гиперплоскости:

$$J_{SFSVM}(a_1, \dots, a_m, b, \delta_1, \dots, \delta_N, C, \mu) = \sum_{i=1}^m q(a_i | \mu) + C \sum_{j=1}^N \delta_j \rightarrow \min(\mathbf{a}, b, \delta),$$

$$q(a_i | \mu) = \begin{cases} 2\mu |a_i| & \text{if } |a_i| \leq \mu, \\ \mu^2 + a_i^2 & \text{if } |a_i| > \mu, \end{cases}$$

$$y_j (\mathbf{a}^T \mathbf{x}_j + b) \geq 1 - \delta_j, \quad \delta_j \geq 0, \quad j=1, \dots, N.$$

μ - степень селективности критерия при выборе информативных отсчетов

$\mu = 0 \rightarrow$ SVM
(отбора признаков нет)
при больших $\mu \rightarrow$ Lasso SVM

Двойственная форма:

$$L(\lambda_1, \dots, \lambda_N | C, \mu) = \sum_{j=1}^N \lambda_j - \sum_{i \in I} \frac{\xi_i}{2} \rightarrow \max(\lambda_1, \dots, \lambda_N),$$

$$\xi_i \geq 0, \quad \xi_i \geq \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \lambda_j \lambda_l - \mu^2, \quad i \in I = \{1, \dots, m\},$$

$$\sum_{j=1}^N y_j \lambda_j = 0, \quad 0 \leq \lambda_j \leq \frac{C}{2}, \quad j=1, \dots, N,$$

* Данный способ регуляризации впервые предложен в работе

Tatarchuk, A., Urlov, E., Mottl, V., Windridge, D.: A support kernel machine for supervised selective combining of diverse pattern-recognition modalities. In: El Gayar, N., Kittler, J., Roli, F. (eds.) MCS (2010)

SVM для обнаружения маммограмм с патологией с отбором наиболее информативных отсчетов ЭЭГ*

Решение задачи нахождения оптимальной разделяющей гиперплоскости:

$$\left\{ \begin{array}{ll} \hat{a}_i = 0, & i \in I^- = \left\{ i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l < \mu^2 \right\}, \text{ не участвуют} \\ \hat{a}_i = \hat{\eta}_i \sum_{j: \hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_{ij}, & i \in I^0 = \left\{ i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l = \mu^2 \right\}, \text{ участвуют с весом 1} \\ \hat{a}_i = \sum_{j: \hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_{ij}, & i \in I^+ = \left\{ i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l > \mu^2 \right\}. \text{ участвуют с весами} \\ & 0 < \eta_i < 1 \end{array} \right.$$

$\hat{\lambda}_j \geq 0, j = 1, \dots, N$ - множители Лагранжа получаемые в результате решения двойственной задачи

Только отсчеты (признаки) из $I^+ \cup I^0$ определяют решающее правило:

$$\sum_{j: \hat{\lambda}_j > 0} y_j \hat{\lambda}_j \left(\sum_{i \in I^+} x_{ij} x_i + \sum_{i \in I^0} \eta_i x_{ij} x_i \right) + b \geq 0,$$

Веса $\eta_i, i \in I^+$ определяются путем решения СЛАУ

* Данный способ регуляризации впервые предложен в работе

Tatarchuk, A., Urlov, E., Mottl, V., Windridge, D.: A support kernel machine for supervised selective combining of diverse pattern-recognition modalities. In: El Gayar, N., Kittler, J., Roli, F. (eds.) MCS (2010)

Качество распознавания маммограмм (AUC)

Обучение: целевых - 98, не целевых - 98

Контроль: целевых - 279, не целевых - 279

Без отбора признаков

	сглаживание	-	-	+	+
	требование гладкости решающего правила	-	+	-	+
номер электрода	27	0,748	0,808	0,728	0,829
	28	0,738	0,830	0,711	0,835
	30	0,772	0,818	0,744	0,825
	32	0,666	0,681	0,644	0,725
	33	0,765	0,831	0,731	0,852
	37	0,720	0,826	0,736	0,845
	39	0,706	0,693	0,681	0,743
	42	0,771	0,858	0,781	0,867
	46	0,643	0,745	0,695	0,767
	53	0,694	0,769	0,729	0,786
	58	0,761	0,714	0,675	0,735
	60	0,749	0,723	0,685	0,747
	63	0,733	0,747	0,701	0,745
	все	0,764	0,805	0,815	0,811

Качество распознавания маммограмм (AUC)

Обучение: целевых - 98, не целевых - 98

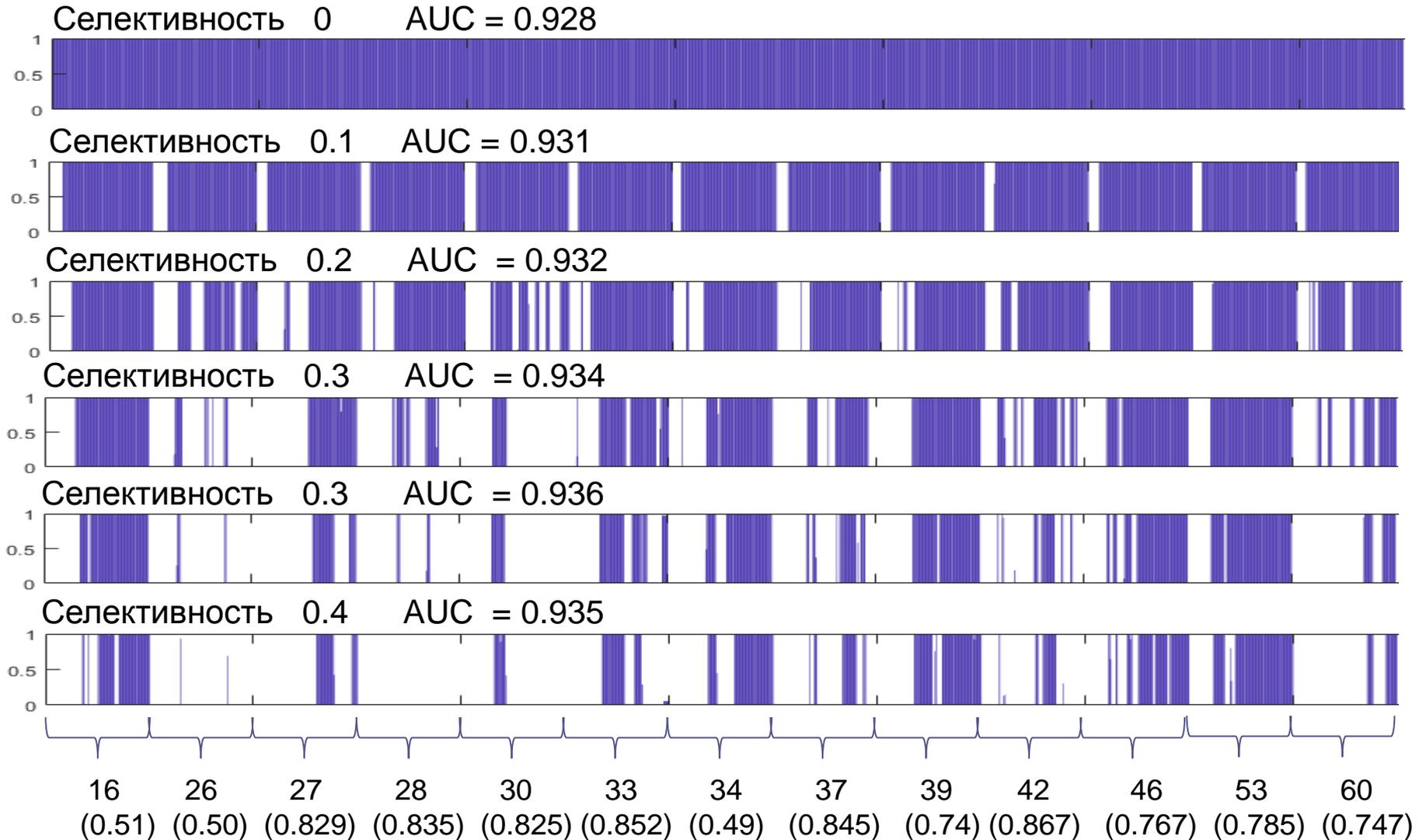
Контроль: целевых - 279, не целевых - 279

Без отбора признаков

	сглаживание	-	-	+	+
	требование гладкости решающего правила	-	+	-	+
номер электрода	27	0,748	0,808	0,728	0,829
	28	0,738	0,830	0,711	0,835
	30	0,772	0,818	0,744	0,825
	32	0,666	0,681	0,644	0,725
	33	0,765	0,831	0,731	0,852
	37	0,720	0,826	0,736	0,845
	39	0,706	0,693	0,681	0,743
	42	0,771	0,858	0,781	0,867
	46	0,643	0,745	0,695	0,767
	53	0,694	0,769	0,729	0,786
	58	0,761	0,714	0,675	0,735
	60	0,749	0,723	0,685	0,747
	63	0,733	0,747	0,701	0,745
	все	0,764	0,805	0,815	0,811

7 электродов, лучших по LOO
AUC = 0.9026

Отобранные признаки для разных значений селективности



Качество распознавания маммограмм (AUC)

Обучение: целевых - 98, не целевых - 98

Контроль: целевых - 279, не целевых - 279

Без отбора признаков

	сглаживание	-	-	+	+
	требование гладкости решающего правила	-	+	-	+
номер электрода	27	0,748	0,808	0,728	0,829
	28	0,738	0,830	0,711	0,835
	30	0,772	0,818	0,744	0,825
	32	0,666	0,681	0,644	0,725
	33	0,765	0,831	0,731	0,852
	37	0,720	0,826	0,736	0,845
	39	0,706	0,693	0,681	0,743
	42	0,771	0,858	0,781	0,867
	46	0,643	0,745	0,695	0,767
	53	0,694	0,769	0,729	0,786
	58	0,761	0,714	0,675	0,735
	60	0,749	0,723	0,685	0,747
	63	0,733	0,747	0,701	0,745
	все	0,764	0,805	0,815	0,811

7 электродов, лучших по LOO
AUC = 0.9026

13 электродов
с отбором признаков
AUC = 0.936

Благодарю за внимание!