



КТ COVID

- CVL (ООО "СиВижинЛаб")
- Гамма Мультивокс Ковирус (ООО "Гаммамед-Софт")
- COVID-19 Intelligent Radiology Assistant (ООО "АЙРА Лабс")



РГ/ФЛГ

- Цельс (ООО "Медицинские скрининг системы")
- Care Mentor AI (ООО "КэреМенторЭйАй")
- FBM (ООО "ФтизисБиоМед")

Комплексная оценка параметров диагностической ценности и качества работы ИИ-Сервиса:

- результаты сравнения с протоколами врачей
- согласие врача с локализацией патологических находок
- удельный вес исследований без дефектов
- выполнение SLA 6.5 минут



КТ COVID

- **CVL**
(ООО "СиВижинЛаб")
- **Care Mentor AI**
(ООО "КэреМенторЭйАй")
- **COVID-19 Intelligent Radiology Assistant** (ООО "АЙРА Лабс")
- **Гамма Мультивокс Ковирус**
(ООО "Гаммамед-Софт")



РГ/ФЛГ

- **Care Mentor AI**
(ООО "КэреМенторЭйАй")
- **FBM**
(ООО "ФтизисБиоМед")

Комплексная оценка параметров диагностической ценности и качества работы ИИ-Сервиса:

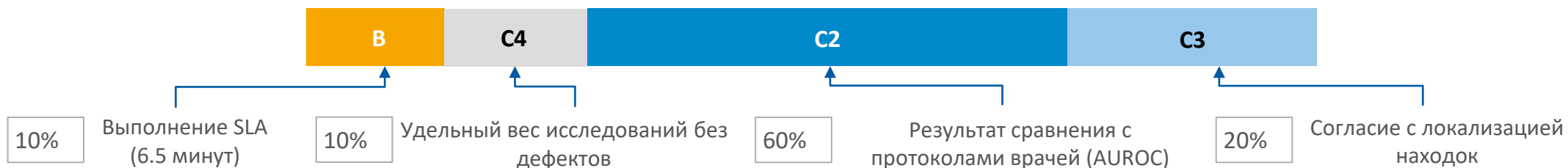
- результаты сравнения с протоколами врачей
- согласие врача с локализацией патологических находок
- удельный вес исследований без дефектов
- выполнение SLA 6.5 минут

* включены направления и сервисы, превысившие пороговый балл

Методология оценки ИИ-Сервисов



Параметр	Описание	Формула расчета	Объем выборки	Диапазон значений	Минимальное значение*	Обоснование минимального значения*	Экспертная оценка**		
							Удельный вес	Обоснование удельного веса	
B	Выполнение SLA (порог 6,5 минут)	$(6,5-t)*15,39$ 0 при $t \geq 6.5$ минут 100 при $t \rightarrow 0$	Вся за отчетный период	[0, 100]	70	Усредненное значение по всем ИИ-Сервисам за первую половину 2021 года по данным выгрузки по форме 3.1	10%	Параметр объективен.	Характеризует техническую надежность и качество работы ИИ-сервиса.
C4	Удельный вес исследований без дефектов «в»-«е»	100 - уд. вес дефектов	Выборка из 20 исследований за отчетный период	[0, 100]	90	НПА Эксперимента	10%		
C2	Минимальная диагностическая точность (AUROC) по результатам проспективного сравнения с текстами протоколов согласно базовым диагностическим требованиям	$100*(AUROC - DI_AUROC/2)$	Вся за отчетный период	(50, 100]	81	Методические рекомендации №43	60%	Параметр объективен.	Характеризует диагностическую точность ИИ-сервиса.
C3 для модальности КТ	Удельный вес согласия врачей, осуществляющих мониторинг, с локализацией находок	Удельный вес согласия врача с локализацией	Выборка из 20 исследований за отчетный период	[0, 100]	70	Экспертная оценка**	20%	Параметр субъективен.	
C3 для модальностей РГ/ФЛГ/ММГ					0				



*Пороговые значения: для включения ИИ-сервиса в рейтинг предусмотрены пороговые значения для итогового балла, соответствующие минимально достаточному уровню качества работы для конечного пользователя ИИ-Сервиса

**по результатам оценки 4 экспертами НПКЦ, удельный вес получен на основании методов ранжирования факторов и парного сравнения (Коробов В. Б. Сравнительный анализ методов определения весовых коэффициентов «влияющих факторов» //Социология: методология, методы и математическое моделирование (Социология: 4М). – 2005. – №. 20. – С. 54-73.)

Методика сравнения ИИ-Сервисов



Цель: объективное сравнение ИИ-сервисов по основным качественным показателям для составления рейтинга.

Методология: сравнение проводится путем расчета для каждого ИИ-сервиса итогового балла за отчетный период. Для этого разработана формула расчета итогового балла, состоящая из 4 параметров и весовых коэффициентов. Весовые коэффициенты в формуле определены методом ранжирования факторов и парного сравнения экспертных оценок*. При определении весовых коэффициентов в формуле, учтена существенно более низкая статистическая значимость параметров, определяемых на ограниченной выборке.

Критерии отбора параметров для формулы:

- параметр характеризует диагностическую точность и/или техническую надежность и/или качество работы ИИ-сервиса;
- возможность получения объективных исходных данных для расчета значения параметра;
- клиническая значимость параметра.

В соответствии с критериями отобраны параметры для формулы:

Параметры **В** и **С4** характеризуют техническую надежность и качество работы ИИ-сервиса (суммарный вклад 20% в итоговый балл).

Параметры **С2** и **С3** характеризуют диагностическую точность ИИ-сервиса (суммарный вклад 80% в итоговый балл).

Параметр	Описание	Формула расчета	Объем выборки	Диапазон значений	Оценивает
В	Выполнение SLA (порог 6,5 минут)	$(6,5-t)*15,39$ 0 при $t \geq 6.5$ минут 100 при $t \rightarrow 0$	Вся за отчетный период	[0, 100]	Группа инженеров и экспертов НПКЦ
С2	Минимальная диагностическая точность (AUROC) по результатам проспективного сравнения с текстами протоколов согласно базовым диагностическим требованиям	$100*(AUROC - ДИ_AUROC/2)$		(50, 100]	
С3	Удельный вес согласия врачей, осуществляющих мониторинг, с локализацией находок	Удельный вес согласия эксперта с локализацией	Выборка из 20 исследований за отчетный период	[0, 100]	Специально обученные врачи-рентгенологи, проводящие ежемесячные мониторинги работы ИИ-Сервисов
С4	Удельный вес исследований без дефектов «в»-«е»	100 - уд. вес дефектов		[0, 100]	

Методика сравнения ИИ-Сервисов



Формула для расчета итогового балла для данного ИИ-сервиса за отчетный период имеет следующий вид:

$$\text{Итоговый балл (D)} = B * 0,10 + C4 * 0,10 + C2 * 0,60 + C3 * 0,20$$

Параметр В – длительность обработки исследования (т.н. «SLA»). Источник: выгрузка по форме 3.1.

Параметр С2 – диагностическая точность (площадь под характеристической кривой – «AUROC»). Источник: проспективный анализ выборки, включающей все результаты работы ИИ-сервиса за отчетный период; референс: протоколы описаний результатов исследований, составленные врачами-рентгенологами МО ДЗМ.

Параметр С3 – согласованность решений о локализации находок. Источник: выборка (из 20) исследований для ручной проверки за отчетный период.

Параметр С4 – удельный вес исследований без дефектов «в»-«е». Источник: выборка (из 20) исследований для ручной проверки за отчетный период

Пороговые значения: для включения ИИ-сервиса в рейтинг предусмотрены пороговые значения для итогового балла, соответствующие минимально достаточному уровню качества работы для конечного пользователя ИИ-Сервиса.

Параметр	Минимальное значение	Обоснование минимального значения
В	70	усредненное значение по всем ИИ-Сервисам за первую половину 2021 года по данным выгрузки по форме 3.1
С2	81	методические рекомендации №43
С3 для модальности КТ	70	экспертная оценка
С3 для модальностей РГ/ФЛГ/ММГ	0	
С4	90	НПА Эксперимента

Установлены следующие значения **пороговых значений итогового балла:**

- Для направления КТ COVID: >80
- Для остальных направлений: >65

Методика сравнения ИИ-Сервисов: пример расчета итоговой оценки ИИ-Сервиса



ИИ-Сервис А:

За отчетный период обработал n исследований КТ ОГК.

- Длительность обработки исследования – 3 минуты (медианное значение по всем n -исследованиям). Показатель $B = (6,5-3)*15,36 = \sim 53,8$
- При расчете диагностической точности:

№ исследования	Врачебная оценка		Ответ ИИ-Сервиса
	Текст заключения	Бинарная оценка	Вероятность патологии
1	КТ-0	0	5
2	КТ-4	1	82
...			
$n-1$	КТ-0	0	10
n	КТ-1	1	51

В результате ROC-анализа, проведенного по всем n -исследованиям, на основании данных из столбца «бинарная оценка» и «вероятность патологии», рассчитано значение площади под характеристической кривой и 95% доверительный интервал: $AUC = 0,91 (0,88-0,94)$. Показатель $C2 = 100*(0,91 - (0,06/2)) = 88$

По результатам выборочной проверки 20 исследований* в ходе ежемесячного мониторинга врачом-Экспертом выявлено:

- 2 исследования с технологическими дефектами. Параметр $C4 = 100*(1 - (2/20)) = 90$
- 14 исследований с целевой патологией, из которых 10 исследований были с корректной локализацией патологических находок. Параметр $C3 = 100*(10/14) = \sim 71,4$

$$\text{Итоговый балл } (D) = 53,8 * 0,10 + 90 * 0,10 + 88 * 0,60 + 71,4 * 0,20 = 82$$

При сравнении ИИ-Сервиса А и ИИ-Сервиса Б, первую строчку в рейтинге занимает ИИ-Сервис А.

ИИ-Сервис А включается в Лидерборд, т.к. превысил пороговое значение по итоговому баллу (для КТ Ковид >80)

ИИ-Сервис Б:

За отчетный период обработал m исследований КТ ОГК.

- Длительность обработки исследования – 2,1 минуты (медианное значение по всем m -исследованиям). Показатель $B = (6,5-2,1)*15,36 = \sim 67,6$
- При расчете диагностической точности:

№ исследования	Врачебная оценка		Ответ ИИ-Сервиса
	Текст заключения	Бинарная оценка	Вероятность патологии
1	КТ-1	1	55
2	КТ-2	1	63
...			
$m-1$	КТ-0	0	21
m	КТ-3	1	89

В результате ROC-анализа, проведенного по всем m -исследованиям, на основании данных из столбца «бинарная оценка» и «вероятность патологии», рассчитано значение площади под характеристической кривой и 95% доверительный интервал: $AUC = 0,83 (0,81-0,85)$. Показатель $C2 = 100*(0,83 - (0,04/2)) = 81$

По результатам выборочной проверки 20 исследований* в ходе ежемесячного мониторинга было выявлено:

- 1 исследование с технологическими дефектами. Параметр $C4 = 100*(1 - (1/20)) = 95$
- 12 исследований с целевой патологией, из которых 6 исследований были с корректной локализацией патологических находок. Параметр $C3 = 100*(6/12) = 50$

$$\text{Итоговый балл } (D) = 67,6 * 0,10 + 95 * 0,10 + 81 * 0,60 + 50 * 0,20 = 75$$

*При формировании случайной выборки из 20 исследований учитывается условие: 75% исследований должны содержать патологию, согласно ответу ИИ-Сервиса