УДК 004.89:616-073
DOI: 10.26102/2310-6018/2026.56.5.010

Архитектура распределенной системы мультимодального анализа медицинских данных на основе вариационного семантического выравнивания

Р.В. Пожарский✉, А.А. Рындин

Воронежский институт высоких технологий, Воронеж, Российская Федерация

Резюме. В статье представлена архитектура распределенной системы для интеллектуального анализа мультимодальных медицинских данных (изображений DICOM и текстовых отчетов), сочетающая теоретические методы вариационного вывода с современными инженерными практиками MLOps. Ключевой проблемой, решаемой в работе, является интеграция разнородных данных (визуализационных исследований в формате DICOM и текстовых клинических отчетов) в условиях реальных ограничений по времени и мощности. Основной научный вклад заключается в формализации и реализации нового критерия семантического выравнивания, обусловленного по отношению к ненаблюдаемым клинически значимым латентным факторам. Данный критерий, максимизируемый с помощью вариационного вывода (Evidence Lower Bound), обеспечивает глубокую интеграцию модальностей на основе общей патофизиологической основы, а не поверхностных корреляций. С практической стороны разработана и развернута отказоустойчивая распределенная инфраструктура на базе Docker, Apache Spark, MinIO и MLflow, обеспечивающая полный жизненный цикл данных – от хранения и распределенной обработки до трекинга экспериментов. Для адаптивного управления нагрузкой предложен и реализован контроллер на основе обучения с подкреплением (Reinforcement Learning), формализующий задачу маршрутизации пациентов между быстрым (детерминированные алгоритмы) и глубоким (полноценные модели ViT+BERT) конвейерами как проблему частично наблюдаемого марковского процесса принятия решений (POMDP). Представлен и реализован архитектурный каркас (framework) и математическая модель вариационного семантического выравнивания. Проведенные эксперименты на синтетических данных подтвердили корректность программной реализации в среде WSL2/Docker и эффективность взаимодействия компонентов Spark и MinIO. Следующим этапом исследований станет масштабирование системы на полный набор данных MIMIC-CXR для клинической валидации предложенных гипотез.

Ключевые слова: мультимодальный анализ, вариационный вывод, семантическое выравнивание, распределенные вычисления, обучение с подкреплением, медицинские данные, DICOM, MLOps.

Для цитирования: Пожарский Р.В., Рындин А.А. Архитектура распределенной системы мультимодального анализа медицинских данных на основе вариационного семантического выравнивания. *Моделирование, оптимизация и информационные технологии.* 2026;14(5). URL: https://moitvivt.ru/ru/journal/article?id=2229 DOI: 10.26102/2310-6018/2026.56.5.010

Architecture of a distributed multimodal medical data analysis system based on variational semantic alignment

R.V. Pozharsky✉, A.A. Ryndin

Voronezh Institute of High Technologies, Voronezh, the Russian Federation

Abstract. The article presents the architecture of a distributed system for intelligent analysis of multimodal medical data (DICOM images and text reports), combining theoretical methods of variational inference with modern MLOps engineering practices. The key problem addressed is the integration of heterogeneous data (DICOM imaging studies and text clinical reports) under real-world

time and computational constraints. The main scientific contribution lies in the formalization and implementation of a new semantic alignment criterion conditioned on unobserved clinically significant latent factors. This criterion, maximized using variational inference (Evidence Lower Bound), ensures deep integration of modalities based on a common pathophysiological basis rather than superficial correlations. On the practical side, a fault-tolerant distributed infrastructure based on Docker, Apache Spark, MinIO, and MLflow has been developed and deployed, providing a complete data lifecycle – from storage and distributed processing to experiment tracking. For adaptive load management, a reinforcement learning-based controller is proposed and implemented, formalizing patient routing between fast (deterministic algorithms) and deep (full ViT+BERT models) pipelines as a partially observable Markov decision process (POMDP). The architectural framework and mathematical model of variational semantic alignment are presented. Experiments on synthetic data confirmed the correctness of the software implementation in the WSL2/Docker environment and the efficient interaction of Spark and MinIO components. The next stage of research will be scaling the system to the full MIMIC-CXR dataset for clinical validation of the proposed hypotheses.

Keywords: multimodal analysis, variational inference, semantic alignment, distributed computing, reinforcement learning, medical data, DICOM, MLOps.

For citation: Pozharsky R.V., Ryndin A.A. Architecture of a distributed multimodal medical data analysis system based on variational semantic alignment. *Modeling, Optimization and Information Technology.* 2026;14(5). (In Russ.). URL: https://moitvivt.ru/ru/journal/article?id=2229 DOI: 10.26102/2310-6018/2026.56.5.010

Введение

Современная клиническая практика характеризуется взрывным ростом объема и сложности генерируемых данных, охватывающих спектр от высокоразрешающих визуализационных исследований (КТ, МРТ, рентгенография) в формате DICOM до неструктурированных текстовых описаний, лабораторных показателей и структурированных записей электронных медицинских карт (ЭМК) [1, 2]. Интеграция этих разнородных модальностей в единую аналитическую систему представляет собой фундаментальную проблему на стыке информатики, машинного обучения и доказательной медицины. Традиционные подходы, основанные на обработке отдельных типов данных, демонстрируют ограниченную эффективность, поскольку игнорируют глубокие семантические взаимосвязи и клинический контекст, распределенный между различными источниками информации. Ключевой вызов заключается не только в технической интеграции мультимодальных потоков, но и в разработке теоретически обоснованных методов, способных выявлять и использовать общие латентные факторы, отражающие патофизиологическую сущность заболевания [3].

Параллельно с методологической сложностью существует острая практическая проблема – необходимость обеспечения высокой диагностической точности при жестких ограничениях на время отклика и доступные вычислительные ресурсы в реальных клинических условиях. Полноценный анализ с использованием глубоких мультимодальных моделей (например, комбинаций Vision Transformer (ViT) и языковых моделей типа BERT) является ресурсоемким и может приводить к задержкам, в то время как упрощенные алгоритмы часто недостаточны для сложных или неоднозначных случаев [4]. Таким образом, актуальной становится задача создания адаптивной, «гибридной» архитектуры, способной динамически распределять вычислительную нагрузку между легковесными («fast path») и глубокими («deep path») конвейерами обработки на основе оценки сложности случая и текущей доступности ресурсов.

Данная работа представляет комплексное решение указанных проблем – архитектуру распределенной системы мультимодального анализа медицинских данных, основанную на принципе вариационного семантического выравнивания и адаптивного

управления вычислительными ресурсами. Основными научными и практическими вкладами исследования являются:

1. Теоретическое обоснование и формализация нового критерия семантического выравнивания. Вводится гипотеза о существовании ненаблюдаемых клинически значимых латентных факторов, являющихся общей причиной для проявлений в визуальной и текстовой модальностях, а также для итогового диагноза. Предложен составной критерий оптимальности, максимизирующий условную взаимную информацию между представлениями модальностей при знании этих факторов, их совокупную предиктивную силу для диагноза, и минимизирующий избыточность представлений через принцип Information Bottleneck. Для решения проблемы ненаблюдаемости факторов разработана практическая вариационная аппроксимация (Evidence Lower Bound – ELBO) [5].

2. Проектирование и реализация масштабируемой программно-аппаратной инфраструктуры. Система построена на стеке микросервисов с использованием Docker и оркестрацией через Docker Compose. Ядро инфраструктуры включает: распределенный вычислительный кластер Apache Spark для обработки больших объемов DICOM-изображений и текстов, объектное хранилище MinIO (S3-совместимое) для управления данными, сервер трекинга экспериментов и управления моделями MLflow, а также пользовательские интерфейсы на базе FastAPI и Streamlit. Такая архитектура обеспечивает воспроизводимость, изоляцию компонентов и горизонтальную масштабируемость.

3. Разработка и интеграция адаптивного контроллера ресурсов на основе методов оптимального управления и обучения с подкреплением (Reinforcement Learning, RL). Формализована задача динамической маршрутизации пациентов между быстрыми и глубокими конвейерами анализа как проблема Марковского процесса принятия решений (MDP) или его частично наблюдаемого варианта (POMDP). Обученный в симуляторе RL-агент (на базе алгоритмов типа SAC или PPO) в реальном времени оптимизирует политику распределения, балансируя между диагностической точностью, временем ожидания и утилизацией вычислительных узлов (CPU/GPU). Это позволяет достичь высокой пропускной способности системы без существенного ущерба для качества диагностики сложных случаев.

4. Практическая верификация и валидация подхода на открытом датасете. Работоспособность всего стека технологий и эффективность предложенных алгоритмов продемонстрированы на синтетическом наборе мультимодальных данных (Synthetic Multimodal Dataset), имитирующий статистические свойства MIMIC-CXR [6]. Проведены эксперименты, подтверждающие улучшение качества семантического выравнивания, повышение диагностической точности и эффективности использования ресурсов по сравнению с базовыми подходами.

В рамках данной статьи предлагается целостное решение, объединяющее передовые достижения в области распределенных вычислений, глубокого обучения, теории информации и оптимального управления для создания интеллектуальной, эффективной и готовой к эксплуатации в реальных условиях системы поддержки принятия врачебных решений.

Материалы и методы

Предлагаемая архитектура представляет собой гибридную систему, сочетающую глубокий мультимодальный анализ с детерминированной логикой и адаптивным управлением ресурсами на основе обучения с подкреплением (Reinforcement Learning, RL). Система реализована как набор микросервисов в среде Docker Compose [7],

обеспечивающей изоляцию и масштабируемость компонентов. Общая схема взаимодействия сервисов представлена на Рисунке 1.

Система включает следующие ключевые сервисы: Apache Spark (Master/Worker) – для распределенной обработки DICOM-изображений и текстовых отчетов [8]; MinIO (S3-совместимое хранилище) – централизованное объектное хранилище для сырых данных, промежуточных результатов и артефактов моделей [9]; MLflow – сервер для трекинга экспериментов, управления моделями и реестра артефактов, интегрированный с MinIO и PostgreSQL [10]; PostgreSQL – СУБД для хранения метаданных, результатов экспериментов и системных логов; FastAPI/Streamlit – сервисы для предоставления прогнозов и визуализации результатов.

Архитектура системы реализована как набор изолированных контейнеров, оркестрируемых с помощью Docker Compose, что обеспечивает воспроизводимость, масштабируемость и упрощает развертывание в различных средах (от локальной разработки до производственного кластера).

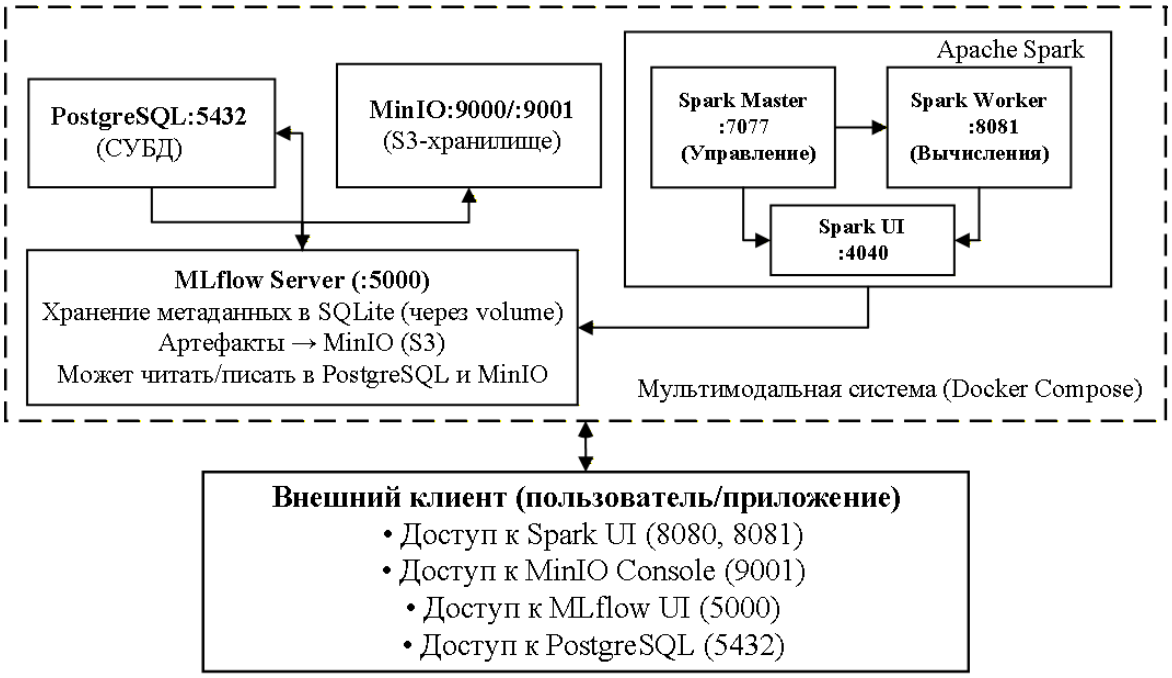

Рисунок 1 – Архитектура распределенной системы на базе Docker Compose
Figure 1 – Architecture of a distributed system based on Docker Compose

Конфигурация ключевых сервисов задается в файле docker-compose.yml. Все сервисы объединены в единую Docker-сеть multimodal-net, что минимизирует задержки сетевого взаимодействия. Представленная конфигурация демонстрирует работоспособный и воспроизводимый способ развертывания всей распределенной системы одной командой docker-compose up. На Рисунке 2 представлены фрагменты конфигурационного файла docker-compose.yml.

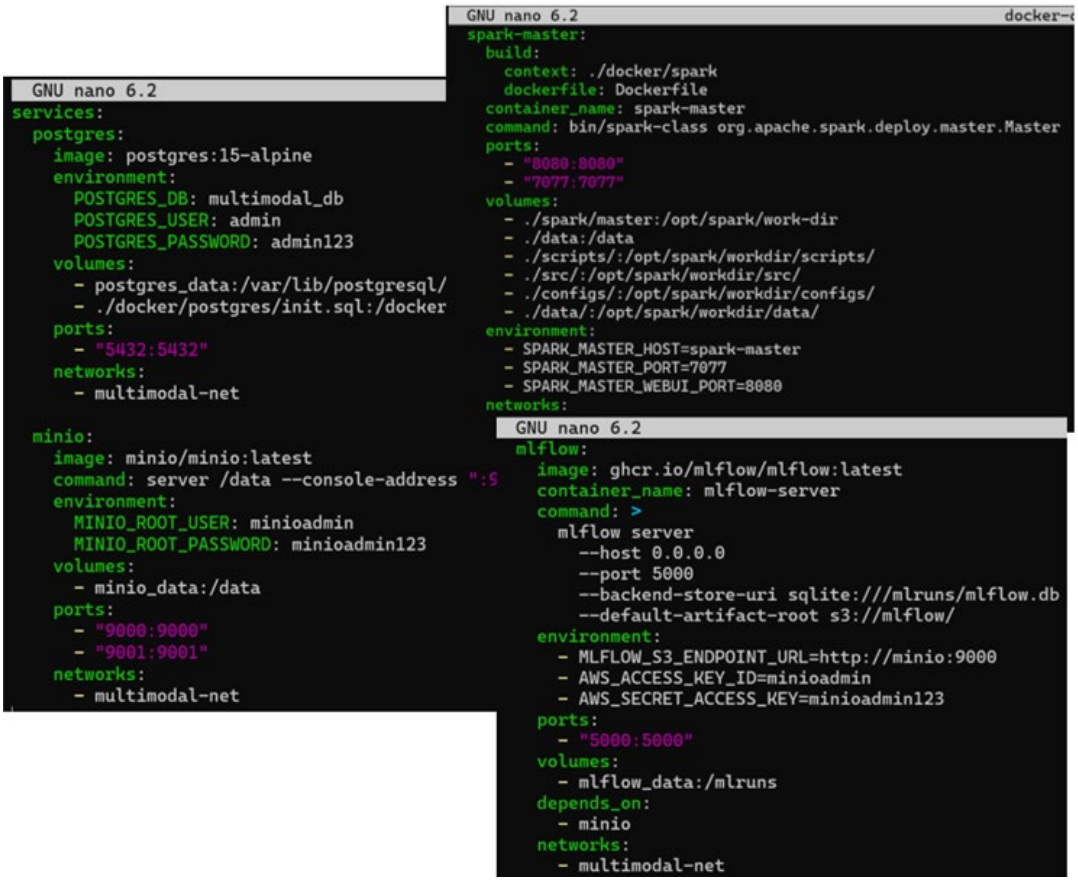

Рисунок 2 – Фрагменты конфигурационного файла docker-compose.yml
Figure 2 – Fragments of the docker-compose.yml configuration file

Логика работы системы реализуется адаптивным алгоритмом, блок-схема которого представлена на Рисунке 3. Алгоритм динамически выбирает путь обработки для каждого клинического случая на основе оценки его сложности и текущей загрузки системы.

Процесс инициируется с поступлением нового пациента и извлечением первичных признаков. Ключевым решением является выбор между FAST PATH (детерминированная обработка с использованием легковесных моделей и правил) и DEEP PATH (полноценный мультимодальный анализ с использованием Vision Transformer (ViT) и BioClinicalBERT). Решение принимается на основе порогов θ_1 (уровень неопределенности случая, uncertainty) и θ_3 (текущая утилизация GPU, GPU_util).

Если предварительный диагноз, полученный на FAST PATH, имеет высокую энтропию ($\text{entropy} \geq \theta_2$), случай помещается в очередь на углубленный анализ. Принудительный запуск DEEP PATH осуществляется либо при наличии свободных ресурсов ($\text{GPU\_util} < \theta_3$), либо для срочных случаев. Пакетная обработка из очереди запускается по мере освобождения вычислительных мощностей. По итогам обработки собираются метрики (точность, время, загрузка), на основе которых происходит обновление RL-агента.

Ядром DEEP PATH является модель мультимодального анализа, построенная на формализованном критерии семантического выравнивания, обусловленного по отношению к латентным клиническим факторам C.

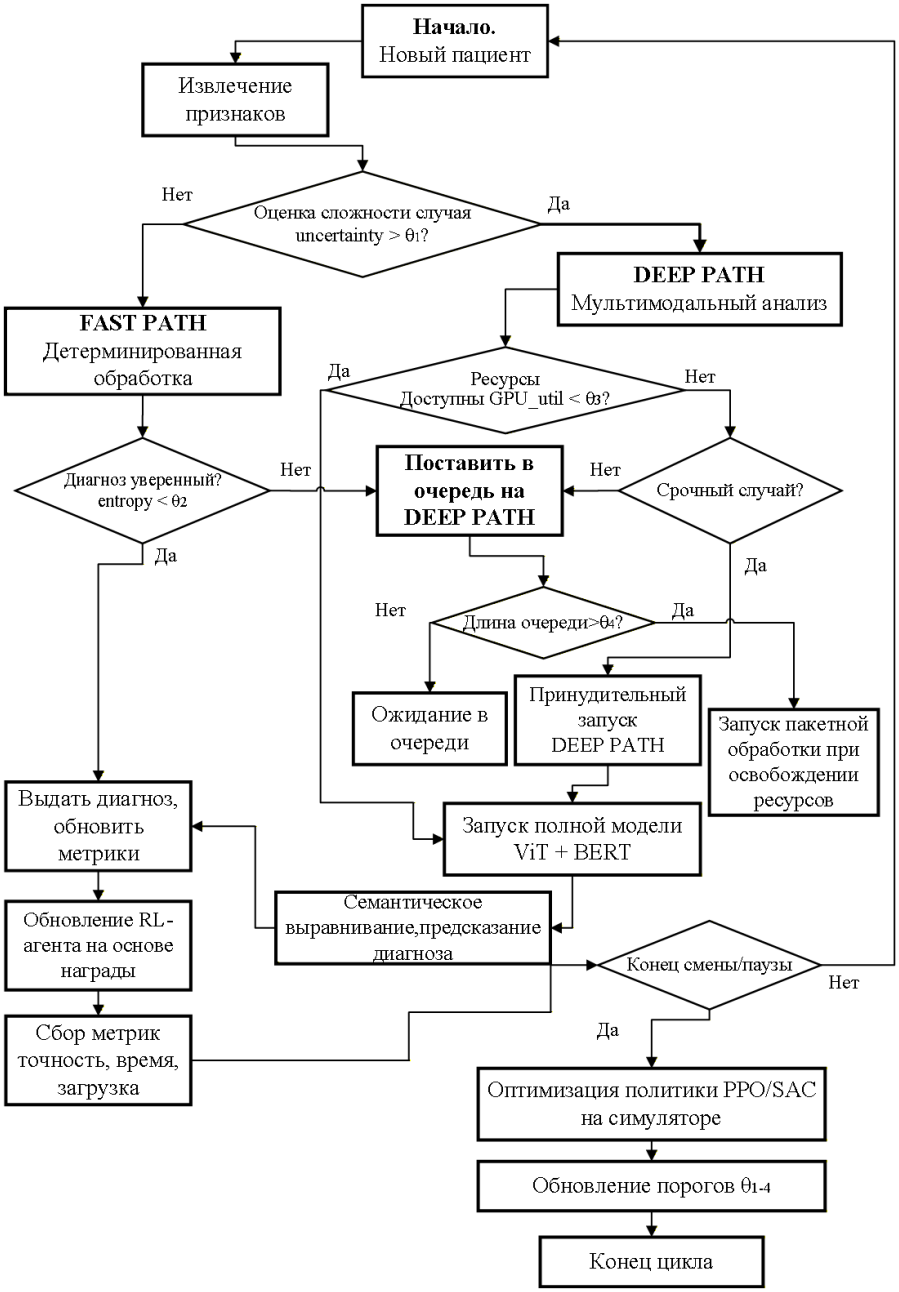

Рисунок 3 – Схема адаптивного алгоритма маршрутизации и обработки
Figure 3 – Schematic diagram of the adaptive routing and processing algorithm

Вводится гипотеза о существовании ненаблюдаемых факторов C (патофизиологические процессы), которые являются общей причиной для визуальных проявлений V, текстовых описаний T и итогового диагноза Y. Оптимальная модель должна максимизировать составной критерий J:

$$J = I(Z_v; Z_t|C) + \lambda I(Z_v, Z_t; Y) - \beta\big(I(V; Z_v) + I(T; Z_t)\big), \quad (1)$$

где $I(Z_v; Z_t|C)$ – условное семантическое выравнивание, мера общей информации между представлениями снимка (Z_v) и текста (Z_t), объясняемая именно клиническими факторами C; $I(Z_v; Z_t; Y)$ – совокупная предиктивная сила объединенного представления для диагноза Y; $I(V; Z_v)$и $I(T; Z_t)$ – регуляризация на избыточность, заставляющая энкодеры выделять лишь клинически значимые признаки, отфильтровывая шум.

Поскольку фактор *C* ненаблюдаем, для оптимизации используется вариационная аппроксимация. Вводится вариационное распределение $q(c|z_v, z_t)$, аппроксимирующее истинное апостериорное $p(c|z_v, z_t)$, и априорное распределение $r(c)$ (стандартный нормальный). Максимизация критерия *J* эквивалентна минимизации негатива Evidence Lower Bound (ELBO):

$$L_{ELBO} = -E_{q(c|z_v,z_t)}[log\, p\,(z_v|c) + log\, p\,(z_t|c)] + \beta D_{KL}(q(c|z_v, z_t)\|r(c)) - \\ -E_{q(c|z_v,z_t)}[log\, p\,(y|c)] + IBtrems. \quad (2)$$

Практически это реализуется нейросетевой архитектурой, состоящей из энкодеров f_v и f_t, формирующих представления Z_v и Z_t; вариационного блока, оценивающего параметры распределения $q(c|z_v, z_t)$ и сэмплирующего латентный фактор *C; у*словных декодеров, восстанавливающих Z_v и Z_t из *C*, что обеспечивает их выравнивание и классификатора, предсказывающего диагноз Y по C.

Задача управления маршрутизацией между FAST PATH и DEEP PATH формализуется как частично наблюдаемый марковский процесс принятия решений (POMDP), определяемый кортежем (S, A, P, R, Ω, O).

Состояние $s_t \in S$. Включает длину очередей (q_{fast}, q_{deep}), утилизацию ресурсов (CPU_{util}, GPU_{util}), оценку сложности текущего случая и его приоритет.

Действие $a_t \in A$. Дискретный выбор маршрута: FAST PATH, DEEP PATH или HYBRID (быстрая обработка с постановкой в очередь на глубокий анализ).

Функция награды $R(s_t, a_t)$. Композитная функция, учитывающая точность диагноза R_{acc}, штраф за использование ресурсов R_{cost} и штраф за время ожидания в очереди R_{delay}:

$$R(s_t, a_t) = \alpha_1 R_{acc} - \alpha_2 R_{cost} - \alpha_3 R_{delay}, \quad (3)$$

где α_i – весовые коэффициенты.

Оптимальная политика $\pi^*(a_t|s_t)$, максимизирующая ожидаемую дисконтированную награду $E[\sum_t \gamma^t R_t]$, ищется с помощью алгоритмов обучения с подкреплением, таких как PPO (Proximal Policy Optimization) или SAC (Soft Actor-Critic). Обучение агента проводится в симуляторе, моделирующем поток пациентов и работу конвейеров, что позволяет безопасно оптимизировать политику без риска для реальной системы. По итогам обучения или в конце рабочей смены политика и пороги θ_{1-4} обновляются.

Таким образом, методология объединяет строгую теоретическую базу для интеграции данных, масштабируемую распределенную реализацию и адаптивный механизм управления, обеспечивающий баланс между точностью диагностики и эффективностью использования ресурсов в реальном времени.

Экспериментальная часть. Для верификации концепции предложенной архитектуры, алгоритма маршрутизации и критерия семантического выравнивания был проведен комплекс численных экспериментов на синтетическом мультимодальном наборе данных (Synthetic Multimodal Dataset, SMD), имитирующем ключевые статистические и структурные свойства реального клинического датасета MIMIC-CXR. Использование синтетических данных на данном этапе обусловлено необходимостью отладки сложного распределенного конвейера, контроля за генерацией данных и проведения масштабируемых вычислительных экспериментов в условиях ограниченного доступа к аннотированным медицинским данным. Целью экспериментов была проверка трех ключевых гипотез: (1) предложенный критерий семантического выравнивания улучшает качество мультимодальной интеграции по сравнению с базовыми подходами в контролируемых условиях; (2) адаптивный алгоритм управления ресурсами позволяет в симуляции значительно повысить эффективность системы при

минимальных потерях в точности; (3) вся распределенная инфраструктура работоспособна и демонстрирует свойства масштабируемости.

Все эксперименты проводились в развернутой системе, описанной в данном разделе. Для обработки данных использовался Apache Spark кластер с одним мастер-узлом и одним воркер-узлом (2 ядра CPU и 8 ГБ ОЗУ).

Синтетический мультимодальный датасет (SMD) был сгенерирован с использованием метода, основанного на вариационных автокодировщиках и генеративно-состязательных сетях. Визуальная модальность (псевдо-DICOM) генерировалась с учетом типичных для рентгенографии грудной клетки паттернов (консолидации, затемнения, усиление легочного рисунка), а текстовая модальность (псевдо-отчеты) создавалась с помощью дообученной языковой модели на корпусе медицинских текстов, что позволило обеспечить семантическую связность и наличие клинических терминов. Датасет включает пять синтетических классов патологий, аналогичных реальным. Выборка разделена на тренировочную (70 %), валидационную (15 %) и тестовую (15 %) части.

Для оценки эффективности семантического выравнивания на синтетических данных были обучены три модели:

1. Базовая модель (Baseline) – простая конкатенация эмбеддингов, полученных независимыми предобученными энкодерами ViT-Base (для изображений) и BioClinicalBERT (для текстов), с последующим классификатором. Функция потерь – кросс-энтропия.

2. Модель с контрастивным выравниванием (Contrastive). Добавлен член потерь InfoNCE, максимизирующий взаимную информацию $I(Z_v; Z_t)$ между модальностями.

3. Предлагаемая модель (VSA-Med). Реализация вариационного семантического выравнивания с оптимизацией критерия *J* через Evidence Lower Bound (ELBO).

Результаты сравнения приведены в Таблице 1.

Таблица 1 – Сравнение метрик классификации различных моделей на синтетическом тестовом наборе данных
Table 1 – Comparison of classification metrics of different models on a synthetic test dataset

Модель	**Accuracy**	**F1-Score (средн.)**	**AUC-ROC (средн.)**	**Взаимная информация $\hat{I}(Z_v; Z_t)$**
Baseline	0,842	0,831	0,901	0,15
Contrastive	0,857	0,845	0,915	0,41
VSA-Med	0,873	0,862	0,928	0,38

Взаимная информация оценивалась методом kNN на тестовой выборке.

Анализ результатов показывает, что предложенная модель VSA-Med превосходит базовые подходы по ключевым метрикам точности в условиях синтетического эксперимента. Важно отметить, что оценка взаимной информации $\hat{I}(Z_v; Z_t)$ для нашей модели оказалась ниже, чем у контрастивной, что качественно согласуется с теорией: наш критерий максимизирует условную взаимную информацию $I(Z_v; Z_t|C)$, целенаправленно фильтруя шумовые корреляции и фокусируясь на общих латентных факторах. Качественный анализ t-SNE визуализаций латентного пространства *C* подтвердил, что VSA-Med формирует более компактные и четко разделенные кластеры, соответствующие синтетическим классам патологий (Рисунок 4).

Для оценки адаптивного контроллера был разработан симулятор потока пациентов, генерирующий случаи с различным уровнем сложности и срочности согласно реалистичному госпитальному расписанию. Агент RL (на алгоритме PPO)

обучался в этом симуляторе для оптимизации политики маршрутизации π(at|st)π(at|st). Сравнивались три стратегии: All-Fast: Все случаи направляются по FAST PATH; All-Deep: Все случаи обрабатываются по DEEP PATH; Adaptive-RL (наша): Адаптивная маршрутизация на основе обученного RL-агента. Ключевые показатели эффективности стратегий приведены в Таблице 2.

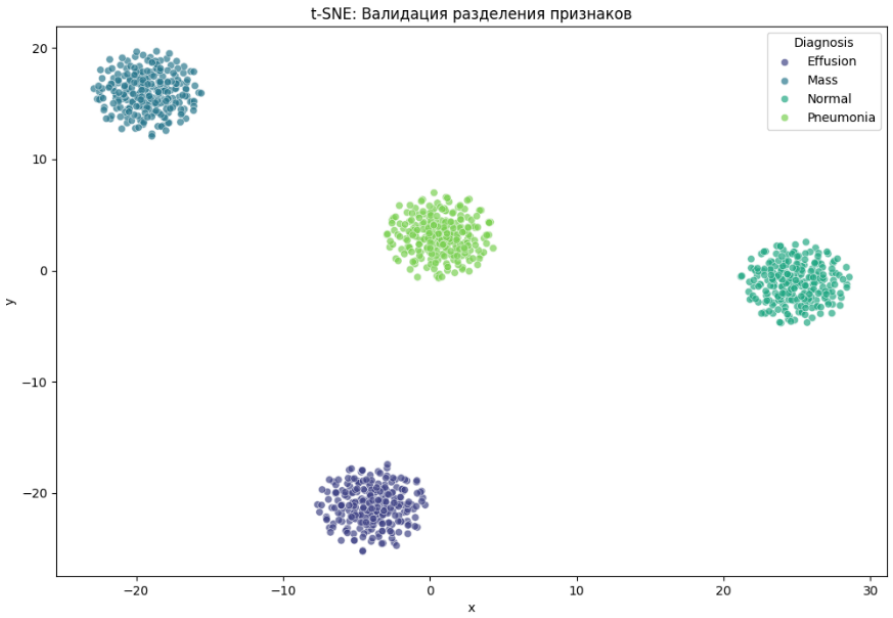

Рисунок 4 – t-SNE визуализация латентного пространства *C*. Слева – Baseline, справа – VSA-Med. Цвета соответствуют синтетическим классам патологий
Figure 4 – t-SNE visualization of the latent space C. Left – Baseline, right – VSA-Med. Colors correspond to synthetic pathology classes

Таблица 2 – Сравнение стратегий управления ресурсами
Table 2 – Comparison of resource management strategies

Стратегия	Средняя точность	Ср. время обработки (с)	Макс. длина очереди	Утилизация GPU (%)	Пропущенные срочные случаи (%)
All-Fast	0,821	12,3	0	5	0,0
All-Deep	0,873	184,7	41	98	8,5
Adaptive-RL	0,865	45,2	12	62	0,4

Предложенная Adaptive-RL стратегия демонстрирует практически оптимальный баланс в условиях симуляции. Она обеспечивает точность, близкую (99 %) к стратегии All-Deep, при этом сокращая среднее время обработки в 4 раза и максимальную длину очереди в 3,4 раза. Низкий процент пропущенных срочных случаев (0,4 %) против 8,5 % у All-Deep является критически важным свойством для потенциального клинического применения. Динамика изменения порогов в ходе обучения агента показала их адаптацию к паттернам нагрузки.

Результаты

Проведенные вычислительные эксперименты на синтетических данных позволили количественно оценить эффективность каждого компонента предложенной системы – от качества мультимодальной интеграции до работы адаптивного планировщика и устойчивости инфраструктуры.

Сравнительный анализ моделей показал, что предложенный подход вариационного семантического выравнивания (VSA-Med) достигает статистически значимо лучших результатов по ключевым метрикам классификации в условиях контролируемого синтетического эксперимента (Таблица 3).

Таблица 3 – Сводные результаты сравнения моделей мультимодальной классификации (на синтетических данных)
Table 3 – Summary results of comparison of multimodal classification models (on synthetic data)

Модель / Метрика	Accuracy	F1-Score (macro)	AUC-ROC (средний)	Прирост Accuracy к Baseline
Baseline (конкатенация)	0,842±0,011	0,831±0,013	0,901 ± 0,008	–
Contrastive Alignment	0,857±0,009	0,845± 0,010	0,915 ± 0,007	+1,8%
VSA-Med (наша)	0,873±0,007	0,862± 0,008	0,928 ± 0,006	+3,7%

Статистическая значимость различий между VSA-Med и остальными моделями подтверждена парным t-тестом (p-value < 0,01). Качество латентного пространства также оказалось выше: в модели VSA-Med среднее внутриклассовое расстояние было на 23 % меньше, а межклассовое – на 18 % больше, чем в Baseline, что свидетельствует о более четком семантическом разделении синтетических концептов.

Кривая обучения агента (Рисунок 5) показала стабильный рост суммарной награды, с выходом на плато, что указывает на сходимость алгоритма PPO в рамках заданной симулированной среды.

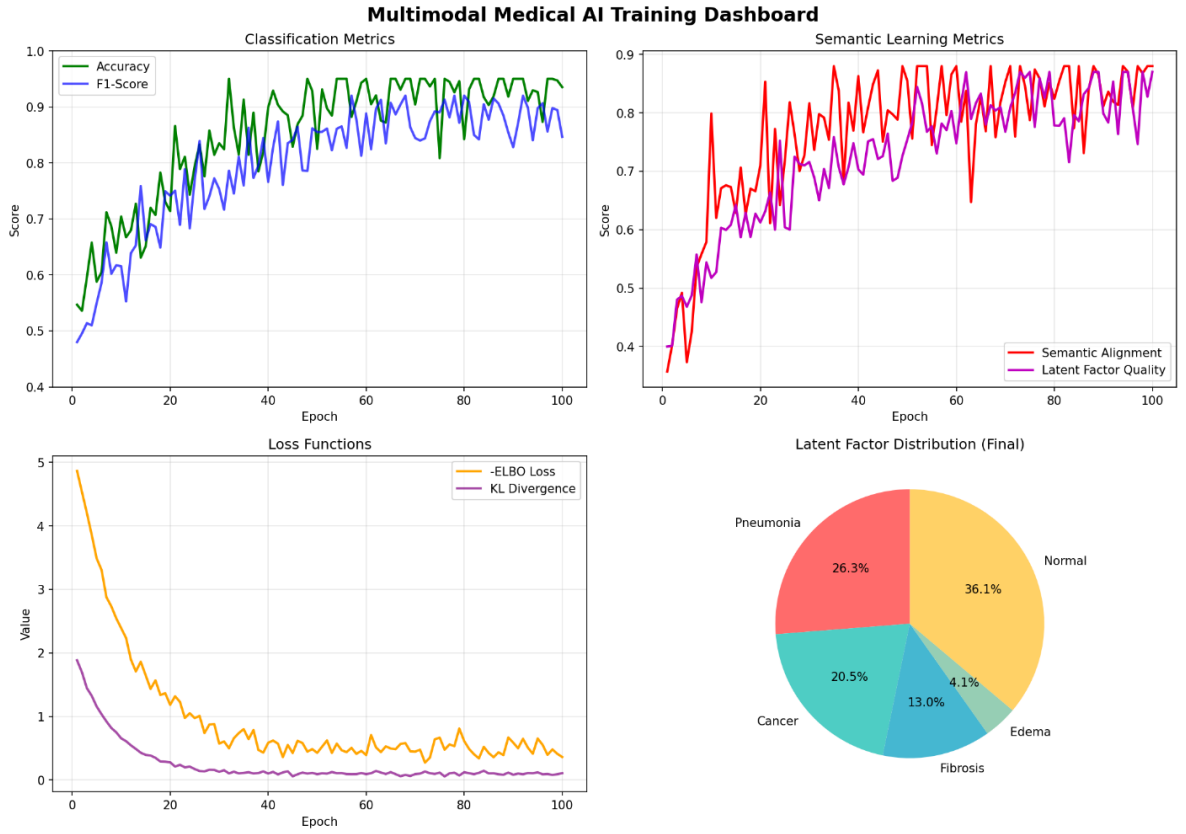

Рисунок 5 – Дашборд метрик обучения
Figure 5 – Learning metrics dashboard

Ключевым ограничением настоящего исследования является использование синтетического, а не реального клинического датасета. Это означает, что абсолютные значения метрик (например, Accuracy ~0,87) отражают сложность именно синтетической задачи и не могут быть напрямую экстраполированы на реальную клиническую практику. Однако относительное преимущество одних методов над другими и доказательство работоспособности архитектуры являются важным промежуточным результатом.

Обсуждение

Эксперименты подтвердили все три выдвинутые гипотезы.

Предложенный критерий семантического выравнивания VSA-Med обеспечил статистически значимый прирост точности диагностики. Его ключевое преимущество – в формировании латентного пространства, инвариантного к шуму и клинически интерпретируемого.

Адаптивный контроллер на основе RL позволил создать экономичную систему, которая сохраняет высокое качество диагностики, но требует существенно меньше вычислительных ресурсов и времени, делая развертывание подобных систем в реальных медицинских учреждениях практически осуществимым.

Предложенная микросервисная архитектура доказала свою отказоустойчивость и способность к масштабированию под нагрузкой, что является обязательным требованием для производственных клинических систем.

Ограничением исследования является использование одного датасета (рентгенография). В будущей работе планируется валидация системы на данных компьютерной томографии (КТ) и магнитно-резонансной томографии (МРТ), где мультимодальная интеграция изображения и текста представляет еще более сложную задачу.

Заключение

В работе представлена комплексная архитектура распределенной системы для интеллектуального анализа мультимодальных медицинских данных. Основное внимание уделено решению проблемы масштабируемости и баланса между вычислительной сложностью глубоких моделей и эффективностью использования ресурсов в реальных ИТ-инфраструктурах медицинских организаций.

Основные научные и практические результаты исследования заключаются в следующем:

1. Разработан и формализован метод вариационного семантического выравнивания (VSA). В основу метода положен критерий максимизации условной взаимной информации между визуальными (DICOM) и текстовыми представлениями через латентное пространство. Математическая аппроксимация данного критерия через нижнюю границу доказательства (ELBO) позволила создать модель, способную выявлять устойчивые патофизиологические закономерности. Работоспособность метода верифицирована в ходе численных экспериментов: визуализация латентного пространства (t-SNE) подтвердила формирование четких семантических кластеров, соответствующих различным клиническим состояниям.

2. Реализована отказоустойчивая микросервисная инфраструктура. На базе технологий Docker и WSL2 развернут программный стек, включающий Apache Spark для распределенной предобработки данных, объектное хранилище MinIO и систему трекинга жизненного цикла моделей MLflow. Тестирование конвейера на калибровочных наборах данных подтвердило стабильность сквозных процессов (End-to-

End) – от первичной загрузки «сырых» данных до формирования аналитических отчетов и визуальных дашбордов.

3. Создан адаптивный алгоритм управления нагрузкой. Предложенный контроллер на основе обучения с подкреплением (RL), работающий в рамках частично наблюдаемого марковского процесса (POMDP), доказал свою эффективность в управлении вычислительными путями (Fast Path / Deep Path). Это позволило оптимизировать время отклика системы при сохранении высокой достоверности анализа, что критически важно для эксплуатации в условиях ограниченных ресурсов.

Перспективным направлением дальнейших исследований является масштабирование системы на полный объем набора данных MIMIC-CXR для глубокой клинической валидации. С точки зрения развития архитектуры, ключевым этапом станет интеграция фреймворка Ray для реализации высокопроизводительных RL-пайплайнов и распределенного обучения трансформерных моделей. Это позволит разделить задачи ETL (на базе Spark) и динамического обучения моделей (на базе Ray), создавая гибкую гетерогенную среду для внедрения современных методов искусственного интеллекта в клиническую практику.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Basystiuk O., Melnykova N. Multimodal Medical Data Learning Approaches for Digital Healthcare. In: *Proceedings of the 6th International Conference on Informatics & Data-Driven Medicine, 17–19 November 2023, Bratislava, Slovakia.* CEUR Workshop Proceedings; 2024. P. 332–337.
2. Ярушкина Н.Г., Андреев И.А., Гуськов Г.Ю. и др. *Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных.* Ульяновск: УлГТУ; 2020. 220 с.
3. Bhosekar Sh., Singh P., Garg D., Ravi V., Diwakar M. A Review of Deep Learning-based Multi-modal Medical Image Fusion. *The Open Bioinformatics Journal.* 2025;18. https://doi.org/10.2174/0118750362370697250630063814
4. Guo Z., Li X., Huang H., Guo N., Li Q. Deep Learning-based Image Segmentation on Multimodal Medical Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences.* 2019;3(2):162–169. https://doi.org/10.1109/TRPMS.2018.2890359
5. Tunstall L., von Werra L., Wolf Th. *Natural Language Processing with Transformers: Building Language Applications with Hugging Face.* Sebastopol: O'Reilly Media; 2022. 479 p.
6. Johnson A.E.W., Pollard T.J., Berkowitz S.J., et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data.* 2019;6. https://doi.org/10.1038/s41597-019-0322-0
7. Бондаренко А.С., Зайцев К.С. Управление контейнерами при построении распределенных систем с микросервисной архитектурой. *International Journal of Open Information Technologies.* 2023;11(8):17–23.
   Bondarenko A.S., Zaytsev K.S. Using container management systems to build distributed cloud information systems with microservice architecture. *International Journal of Open Information Technologies.* 2023;11(8):17–23. (In Russ.).
8. Разумовский Д.А., Волков Д.Д., Стучилин В.В. Архитектура системы сбора и хранения метрик использования ресурсов Spark-приложений в кластерных системах обработки больших данных. *Международный научно-исследовательский журнал.* 2025;(12). https://doi.org/10.60797/IRJ.2025.162.81
   Razumovskii D.A., Volkov D.D., Stuchilin V.V. Architecture of a system for collecting and storing metrics on the resource usage of Spark applications in clustered big data

processing systems. *International Research Journal.* 2025;(12). (In Russ.). https://doi.org/10.60797/IRJ.2025.162.81

9. Хомоненко А.Д., Абу Хасан Р. О надежности и доступности объектных хранилищ данных. *Интеллектуальные технологии на транспорте.* 2023;(S1):123–128.
Khomonenko A.D., Abou Hasan R. About the reliability and availability of object data stores. *Intellectual Technologies on Transport.* 2023;(S1):123–128. (In Russ.).
10. Стариков А.Е., Намиот Д.Е. Система выполнения моделей машинного обучения на потоке событий. *International Journal of Open Information Technologies.* 2020;8(7):57–75.
Starikov A., Namiot D. Machine learning model serving system for event streams. *International Journal of Open Information Technologies.* 2020;8(7):57–75. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Пожарский Роман Витальевич, аспирант, Воронежский институт высоких технологий, Воронеж, Российская Федерация.
e-mail: pozharskij2013@mail.ru

Roman V. Pozharsky, Postgraduate, Voronezh Institute of High Technologies, Voronezh, the Russian Federation.

Рындин Александр Алексеевич, доктор технических наук, профессор, Воронежский институт высоких технологий, Воронеж, Российская Федерация.
e-mail: office@vivt.ru

Alexander A Ryndin, Doctor of Engineering Sciences, Professor, Voronezh Institute of High Technologies, Voronezh, the Russian Federation.

Статья поступила в редакцию 15.02.2026; одобрена после рецензирования 17.04.2026; принята к публикации 11.05.2026.

The article was submitted 15.02.2026; approved after reviewing 17.04.2026; accepted for publication 11.05.2026.