

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теории функций и стохастического анализа

**РАЗРАБОТКА ПРИКЛАДНОГО ПРОГРАММНОГО
ОБЕСПЕЧЕНИЯ ПО БИНАРНОЙ КЛАССИФИКАЦИИ
РЕНТГЕНОВСКИХ СНИМКОВ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы
направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Ежова Андрея Дмитриевича

Научный руководитель

д. ф.-м. н., доцент

С. П. Сидоров

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы. В современной клинической практике и эпидемиологии автоматизация процессов медицинской визуализации представляет собой одно из наиболее приоритетных и быстроразвивающихся направлений прикладной математики и компьютерных наук. Внедрение методов глубокого машинного обучения в сферу лучевой диагностики позволяет существенно повысить пропускную способность медицинских учреждений и снизить процент диагностических ошибок, вызванных человеческим фактором и усталостью в условиях дефицита узкопрофильных специалистов. Создание специализированного прикладного программного обеспечения, интегрирующего в себе передовые нейросетевые модели, является актуальной научно-технической задачей, лежащей на стыке искусственного интеллекта, медицинской информатики и системного программирования.

Цель работы. Требуется разработать модель машинного обучения, способную автоматически классифицировать рентгеновские снимки грудной клетки на два класса:

Класс 1 - Наличие признаков туберкулёза (ТВ).

Класс 2 - Нормальные лёгкие (отсутствие патологий).

Также создать графический интерфейс десктопного приложения для интуитивного использования модели.

Задачи выполняемой ВКР. Для достижения поставленной цели в работе сформулированы и последовательно решены следующие исследовательские и инженерные задачи:

1. Провести анализ данных: изучить количество изображений, оценить баланс классов и определить пропорции разбиения выборки.
2. Реализовать методы аугментации данных для улучшения обобщающей способности модели.
3. Выбрать наиболее подходящие для будущих моделей оптимизатор, функцию потерь и метрики.
4. Разработать три кардинально различные нейросетевые архитектуры для обеспечения разнообразия извлекаемых признаков.
5. Оценить качество моделей с помощью метрик.

6. Проанализировать ошибки классификации.
7. Объединить обученные модели в гетерогенный ансамбль.
8. Оценить качество системы и провести анализ матрицы ошибок с учётом медицинских требований к скринингу.
9. Разработать графическую оболочку для прикладного использования модели.

Краткая характеристика материалов исследования. В ходе исследования использовались фреймворки глубокого обучения TensorFlow/Keras и анализа данных Scikit-learn, а также специализированные библиотеки для обработки информации. Для проектирования графической оболочки применён встроенный кроссплатформенный инструментальный Tkinter. Экспериментальная часть работы базируется на верифицированном открытом наборе данных «Tuberculosis (TB) Chest X-ray Dataset», сформированном консорциумом исследователей Катарского университета, Университета Дакки и профильных медицинских экспертов. Общий объём анализируемой выборки составляет 4200 цифровых рентгенограмм в высоком разрешении. Выборка характеризуется выраженным дисбалансом классов, отражающим реальную эпидемиологическую картину: 3500 снимков представляют физиологическую норму (класс «Normal»), а 700 снимков содержат подтверждённые радиологические признаки туберкулёза (класс «Tuberculosis»). Теоретической базой послужили последние исследования в области компьютерного зрения и медицинской диагностики.

Описание структуры ВКР. Выпускная квалификационная работа изложена на 83 страницах, состоит из введения, четырёх глав, заключения и двух приложений. В приложениях к работе представлены полные листинги программного кода, содержащие реализацию нейросетевых архитектур, процедуру обучения ансамбля и исходный код графического интерфейса.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе «Теоретические основы нейронных сетей для анализа изображений» приводится строгое математическое описание парадигмы машинного обучения с учителем. Обучающая выборка формализуется как множество упорядоченных пар:

$$X^l = \{(x_i, y_i)_{i=1}^l\},$$

где каждый объект x_i представляет собой вектор в n -мерном пространстве признаков \mathbb{R}^n (матрицу пикселей изображения), а $y_i \in \{0, 1\}$ — соответствующую ему истинную бинарную метку целевого класса. Описывается процедура извлечения вторичных информативных признаков с целью перехода в признаковое пространство меньшей размерности:

$$F = \|f_j(x_i)\|_{l \times k}, \quad k \ll n.$$

Цель обучения формулируется как минимизация эмпирического риска — среднего арифметического значения функции потерь $L(a, x)$ на всей обучающей выборке для построения оптимальной аппроксимирующей функции (модели) $a(x)$:

$$Q(a, X^l) = \frac{1}{l} \sum_{i=1}^l L(a, x_i) \rightarrow \min_a$$

В качестве примеров базовых функций потерь рассмотрены абсолютная и квадратичная ошибки. Описывается структура искусственного нейрона и правила функционирования многослойного персептрона, где передача сигнала между слоями описывается линейным преобразованием с последующим применением нелинейной активации: $z = f(W^T x + b)$. Математически разобран алгоритм обратного распространения ошибки, основанный на вычислении частных производных целевой функции по весовым коэффициентам сети с использованием цепного правила дифференцирования сложных функций.

Далее подробно излагается математический аппарат свёрточных нейронных сетей (CNN). Операция дискретной двумерной свёртки входного тен-

зора изображения I с ядром фильтра K размером $m \times n$ записывается в виде:

$$(I * K)(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} I(x+i, y+j) \cdot K(i, j).$$

Описывается влияние параметров шага свёртки и дополнения границ нулями на результирующую пространственную размерность выходной карты признаков. Операция субдискретизации MaxPooling для окна размером 2×2 формализуется как нелинейное отображение, оставляющее максимальный элемент из локального подмассива:

$$\text{MaxPooling}(S)_{X,Y} = \max_{a,b=0}^1 S_{2X+a, 2Y+b},$$

Важнейшей теоретической частью первой главы является детальный вывод разложения среднеквадратичной ошибки предсказания модели на смещение, дисперсию и неустраняемый случайный шум σ_ϵ^2 . Пусть истинная зависимость имеет вид $y(x) = f(x) + \epsilon$, где $E[\epsilon] = 0$ и $V[\epsilon] = \sigma_\epsilon^2$. Ожидаемая квадратичная ошибка модели $a(x)$ в фиксированной точке x преобразуется следующим образом:

$$\text{Err}(x) = \mathbb{E}[(a(x) - y(x))^2] = \mathbb{E}[(a(x) - f(x) + \epsilon)^2].$$

Учитывая взаимную независимость шума и ответов алгоритма, раскрывается квадрат разности:

$$\text{Err}(x) = \text{Bias}^2(\bar{a}) + \mathbb{V}[\bar{a}] + \sigma_\epsilon^2.$$

На основе полученного разложения анализируется поведение среднего арифметического ансамбля из M моделей: $\bar{a}(x) = \frac{1}{M} \sum_{m=1}^M a_m(x)$. Доказывается, что математическое ожидание ансамбля сохраняет среднее смещение его компонентов, в то время как дисперсия ансамбля $\mathbb{V}[\bar{a}]$ при одинаковой дисперсии базовых моделей σ^2 и среднем коэффициенте корреляции их ошибок ρ претерпевает существенное снижение:

$$\mathbb{V}[\bar{a}] = \mathbb{V} \left[\frac{1}{M} \sum_{m=1}^M a_m \right] = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \text{Cov}(a_i, a_j).$$

$$\mathbb{V}[\bar{a}] = \rho\sigma^2 + \frac{1 - \rho}{M}\sigma^2.$$

Этот вывод математически доказывает ключевую гипотезу исследования: минимизация корреляции ρ между ошибками базовых моделей путём построения гетерогенной структуры позволяет радикально снизить итоговую дисперсию предсказаний системы.

Во втором разделе «Реализация ансамбля моделей» описывается пошаговое построение программного комплекса. На этапе предобработки исследована проблема сильного дисбаланса классов (отношение нормы к патологии составляет 5:1). Обоснован выбор бинарной кросс-энтропии в качестве минимизируемой функции потерь, которая для истинного класса y и предсказанной моделью вероятности \hat{y} записывается как:

$$Loss(y, \hat{y}) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y})).$$

Данная функция является строго выпуклой в интервале $(0, 1)$, что гарантирует отсутствие локальных минимумов при градиентном спуске. В качестве оптимизатора выбран алгоритм Adam, сочетающий идеи методов Momentum и RMSProp. Шаг обновления весов ω в момент времени t описывается системой уравнений:

$$\begin{aligned} \nu_t &= \beta_1 \cdot \nu_{t-1} - (1 - \beta_1) \cdot g_t \\ s_t &= \beta_2 \cdot s_{t-1} - (1 - \beta_2) \cdot g_t^2 \\ \Delta\omega_t &= -\lambda \frac{\nu_t}{\sqrt{s_t + \epsilon}} \\ \omega_{t+1} &= \omega_t + \Delta\omega_t, \end{aligned}$$

где λ — скорость обучения, g_t — градиент функции потерь, ν_t и s_t — экспоненциальные скользящие средние первого и второго моментов градиента соответственно, β_1, β_2 — параметры затухания, $\epsilon = 10^{-8}$. Для оценки качества моделей формализованы медицинские метрики на основе матрицы ошибок:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

Метрика Recall (полнота) минимизирует пропуск больных пациентов, а Specificity (специфичность) контролирует ложные тревоги. Также введена интегральная метрика ROC-AUC.

Далее подробно описывается архитектурное проектирование трёх разработанных независимых гетерогенных нейросетевых моделей:

1. *Архитектура на базе плотных свёрточных связей.* Модель построена на концепции сквозного прохождения градиентов, где каждый слой получает на вход конкатенацию карт признаков всех предыдущих слоёв. Скорость роста сети выбрана как $k = 16$. Базовый блок включает в себя последовательное применение пакетной нормализации, функции активации ReLU и свёртки с ядром 3×3 . Пакетная нормализация для мини-пакета $B = \{x_i\}_{i=1}^m$ устраняет эффект внутреннего сдвига ковариат и описывается уравнениями:

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i, \quad \sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad y_i = \gamma \hat{x}_i + \beta$$

Сеть состоит из трёх плотных блоков, разделённых слоями перехода, состоящими из свёртки 1×1 и AveragePooling 2×2 . Выходной вектор формируется через GlobalAveragePooling и полносвязный слой с функцией сигмоиды и регуляризацией Dropout(0.4). При программном изменении порога до 0.3 модель показала высочайшую чувствительность (Recall = 0.9905), но невысокую специфичность (Specificity = 0.7695).

2. *Архитектура на основе многослойных перцептронов.* Данная модель полностью отказывается от свёрточных и аппроксимирующих слоёв. Входная матрица изображения 224×224 разбивается на $S = 196$ непересекающихся патчей размером 16×16 пикселей. Каждый патч линейно проецируется в вектор размерности $C = 64$. Полученный тензор размера 196×64 последовательно обрабатывается блоками смешивания двух типов: пространственного и канального. Пространственное смешивание перемешивает информацию между патчами, оперируя столбцами матрицы, а канальное — внутри каждого отдельного патча, оперируя строками. Внутри блоков используется послойная нормализация и связи со сдвигом. На тестовой выборке модель

проявила себя как консервативный фильтр: $\text{Recall} = 0.9238$, но крайне высокая специфичность $\text{Specificity} = 0.9790$.

3. *Архитектура на базе глубокого свёрточного автоэнкодера.* Модель реализует концепцию обучения признаков без учителя. Энкодер, состоящий из четырёх последовательных слоёв свёртки и MaxPooling, обучался сжимать только изображения класса «Normal» в компактное латентное пространство размерности $28 \times 28 \times 128$. Декодер восстанавливал исходную матрицу, минимизируя ошибку MSE. После этапа сходимости декодер удалялся, веса энкодера полностью замораживались, а к латентному слою подключался классификационный многослойный перцептрон. Обучаясь на бинарную классификацию, эта модель научилась улавливать более абстрактные аномалии на снимках лёгких. Модель показала ультраконсервативное поведение: $\text{Recall} = 0.5472$, а $\text{Specificity} = 0.9942$.

В заключительной части второй главы реализована процедура построения взвешенного ансамбля. Итоговое предсказание формируется как линейная комбинация выходов трёх сетей:

$$\hat{P}_w(y(x) = 1|x) = w_1 \cdot a_1(x) + w_2 \cdot a_2(x) + w_3 \cdot a_3(x).$$

С помощью численного метода поиска по сетке на валидационной выборке были определены оптимальные весовые коэффициенты, минимизирующие ковариацию ошибок: $w_1 = 0.25$, $w_2 = 0.60$, $w_3 = 0.15$. Порог принятия бинарного решения был смещён до уровня $T = 0.25$ для полного исключения пропуска болезни. Итоговые метрики ансамбля на тестовой выборке составили: Полнота (Recall) = 0.9906, Специфичность (Specificity) = 0.8705, ROC-AUC = 0.9925. Объединение позволило взаимно компенсировать дисперсию моделей, выдав оптимальный скрининговый результат.

В третьем разделе «Разработка прикладного программного обеспечения для работы с моделью» подробно рассматривается процесс системного проектирования десктопного программного обеспечения. Логическая структура приложения жёстко следует паттерну проектирования Model-View-Controller (MVC), что гарантирует слабую связанность компонентов и легкость масштабирования программного кода. С точки зрения программной инженерии,

архитектура приложения базируется на классическом паттерне проектирования «Модель-Представление-Контроллер» (MVC), обеспечивающем чёткое разделение данных, интерфейса и управляющей логики:

1. **Модель** инкапсулирует бизнес-логику системы. В её роли выступает предварительно обученный гетерогенный ансамбль, сохранённый в виде единого программного объекта, а также математические алгоритмы предобработки входных тензоров.
2. **Представление** отвечает за визуальное взаимодействие с пользователем. Включает адаптивную область предпросмотра снимков, кнопки управления, цветовые индикаторы уверенности модели и текстовые поля для вывода интерпретируемого диагноза.
3. **Контроллер** обрабатывает события интерфейса и управляет потоком данных. При получении команды от пользователя контроллер извлекает изображение из Представления, передаёт его в Модель для получения предсказания и возвращает вычисленный результат обратно в Представление для отрисовки.

В четвёртом разделе «Сравнение с существующими решениями» осуществлена строгая экспериментальная верификация разработанного подхода. Проведено сравнительное тестирование предложенного гетерогенного ансамбля с популярными глубокими свёрточными сетями — ResNet-50 и VGG-16. Эксперименты показали, что базовая сеть ResNet-50 со стандартным порогом классификации демонстрирует высокую стабильность и хороший баланс общих метрик ($\text{Recall} = 0.9622$, $\text{Specificity} = 0.9657$), однако в условиях реального скрининга показатель пропусков на уровне почти 4% является критическим и эпидемиологически опасным. Разработанный взвешенный гетерогенный ансамбль продемонстрировал $\text{ROC-AUC} = 0.9925$. Благодаря эффекту смешивания признаков различной природы и целенаправленной калибровке порога до 0.25, полнота выявления патологии ансамблем была доведена до уровня 0.9906, что минимизирует ложноотрицательные результаты до единичных случаев на всей генеральной совокупности. Вызванное этим контролируемое снижение специфичности до 0.8705 признано полностью оправданным для задачи выявления всех больных пациентов.

ЗАКЛЮЧЕНИЕ

В рамках выпускной квалификационной работы успешно решена актуальная научно-практическая задача создания автоматизированного программного комплекса для бинарной классификации цифровых рентгенограмм органов грудной клетки на предмет выявления признаков туберкулёза лёгких.

Проведённые теоретические изыскания и численные эксперименты полностью подтвердили исходную математическую гипотезу: объединение глубоких нейросетевых моделей принципиально различной топологии в единый гетерогенный взвешенный ансамбль позволяет минимизировать корреляцию их ошибок и существенно снизить общую дисперсию предсказательной системы на новых, ранее не встречавшихся данных. Разработанные базовые классификаторы (плотная свёрточная сеть, модель смешивания токенов MLP-Mixer и глубокий автоэнкодер, обученный без учителя) продемонстрировали ортогональные профили распределения ошибок, что послужило базисом для их успешной интеграции.

Разработанный алгоритм взвешенного голосования с подобранными коэффициентами (0.25, 0.60, 0.15) и нелинейно смещённым порогом чувствительности до уровня 0.25 позволил достичь выдающихся для систем первичного медицинского скрининга результатов. Итоговый показатель полноты (Recall) составил 0.9906 при сохранении высокого уровня специфичности (Specificity = 0.8705) и интегральной разделяющей способности (ROC-AUC = 0.9925). Разработанный комплекс значительно превзошёл по показателям безопасности классические архитектуры глубокого обучения (ResNet-50, VGG-16), минимизировав риск критического пропуска больного пациента.

Практическая значимость работы заключается в реализации созданного алгоритмического ядра в виде полноценного автономного десктопного приложения с графическим интерфейсом пользователя, спроектированного по паттерну MVC. Разработанное программное обеспечение функционирует локально, не требует дорогостоящих серверных мощностей, автоматически производит полный цикл предобработки растровых матриц и выдаёт мгновенный интерпретируемый результат с цветовой индикацией рисков. Программный продукт готов к внедрению в качестве СППВР в медицинских ка-

бинетах первичного приёма. Перспективы развития проекта связаны с расширением обучающей выборки, переходом к мультиклассовой детекции сопутствующих лёгочных патологий (пневмонии, новообразования) и интеграцией поддержки медицинского формата передачи данных DICOM.