

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования



**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Направление подготовки/профиль 09.06.01 Информатика и вычислительная техника
Инженерная школа информационных технологий и робототехники
Отделение информационных технологий

**Научный доклад об основных результатах подготовленной
научно-квалификационной работы**

Тема научного доклада
Математическое и программное обеспечение для локализации патологий диссеминированной формы туберкулёза лёгких по данным компьютерной томографии
УДК <u>004.891.3:004.415.2:620.179.152.1:616-002.5</u>

Аспирант

Группа	ФИО	Подпись	Дата
А6-39	Костин Кирилл Александрович		

Руководитель профиля подготовки

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ОИТ ИШИТР	Спицын В.Г.	Д.Т.Н., профессор		

Руководитель отделения

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Исполняющий обязанности руководителя ОИТ ИШИТР	Шерстнев В.С.	К.Т.Н., доцент		

Научный руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Аксёнов С.В.	К.Т.Н., доцент		

**АННОТАЦИЯ К НАУЧНОМУ ДОКЛАДУ ОБ ОСНОВНЫХ
РЕЗУЛЬТАТАХ ПОДГОТОВЛЕННОЙ НАУЧНО-
КВАЛИФИКАЦИОННОЙ РАБОТЫ**

**«МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ
ЛОКАЛИЗАЦИИ ПАТОЛОГИЙ ДИССЕМНИРОВАННОЙ ФОРМЫ
ТУБЕРКУЛЁЗА ЛЁГКИХ ПО ДАННЫМ КОМПЬЮТЕРНОЙ ТОМОГРАФИИ»**

Ключевые слова: туберкулёз, компьютерная томография, программное обеспечение, машинное и глубокое обучение, компьютерное зрение.

Туберкулёз лёгких является одним из опаснейших заболеваний, входящих в десятку основных причин смерти человека по всему миру по данным Всемирной организации здравоохранения. Для решения проблем его диагностики всё чаще применяются компьютерные медицинские системы, использующие снимки рентгена и компьютерной томографии (КТ) как наиболее информативные. Одним из направлений развития исследований и разработок в этой области является дифференциальная диагностика различных форм заболеваний и локализация патологий с использованием методов визуализации. На сегодняшний день наблюдается интенсивное развитие систем автоматической сегментации органов и аномальных опухолевых образований, значительно повышающих эффективность принятия врачебных решений. Наибольший вклад в развитие данного направления внесли: Liang J., Lee M., Karargyris A., Jaeger S., Rosenthal A., Ramya R., Ковалёв В.А., Тухбатуллин М.Г., Тюрин И.Е. Несмотря на огромный интерес научного сообщества, проблема диагностики форм туберкулёза лёгких, предполагающая локализацию патологий на снимках КТ, затронута слабо и далека от решения.

В научно-квалификационной работе рассмотрен комплекс методов и программное обеспечение, предлагаемые для решения задачи локализации

патологий диссеминированной формы туберкулёза по данным указанных радиологических исследований.

Для выделения участков лёгочной ткани на срезах предлагается алгоритм сегментации, использующий кластеризацию денситометрических данных и подходы компьютерного зрения для выделения связных областей пикселей, соответствующих регионам, характеризующихся низкой плотностью. Подход выделяет области лёгких в 99,2 % случаев, а реализация алгоритма позволяет эффективно распараллелить его выполнение на множестве процессорных ядер.

Локализация областей интереса, визуально схожих с патологическими, осуществляется с помощью разработанной автором модели сегментации внутрилёгочных структур на основе глубокой свёрточной нейронной сети с энкодер-декодер архитектурой, прототипом которой была архитектура U-net. Для повышения качества работы по локализации регионов интереса осуществлены следующие модификации: в качестве функции ошибки была выбрана функция на основе индекса Сёренсена-Дайса, учитывающая несбалансированность пикселей классов сегментов; добавлены слои нормализации батчей; произведена замена функций активации внутренних слоёв сети, также обучающая выборка расширена с помощью искусственных аугментаций. Разработанная модель допускает малое количество пропусков и обеспечивает качество со значениями метрик: 0,52 – точность, 0,90 – полнота, 0,66 – F-мера.

Для классификации областей интереса с последующей локализацией патологий произведена оценка эффективности моделей логистической регрессии, случайного леса и градиентного бустинга, обученных на выборке текстурных признаков, выделенных с помощью алгоритмов GLCM и LBP. Для различных способов формирования обучающих данных осуществлена оптимизация параметров алгоритмов расчета текстурных признаков и выбрана наиболее эффективная модель классификации: градиентный бустинг над

деревьями решений с максимальной глубиной 7, обученный на текстурных признаках GLCM со значениями гиперпараметров ширины окна 25 и количества уровней квантования серого цвета, равной 5.

Предлагаемые методы реализованы в программной системе, включающей в себя следующие компоненты: платформу для анализа данных и проведения экспериментов, сервер управления данными КТ пациентов и графический интерфейс пользователя, представляющий собой web-приложение. Компонент платформы является основным функциональным ядром и включает в себя блок, обеспечивающий обучение, тестирование и оценку моделей, а также блок, предназначенный для применения обученных моделей в задачах локализации патологий, реализуемых программной системой. Для разработки были использованы языки программирования Python и JavaScript, технологии Django и React, а также инструменты для машинного и глубокого обучения TensorFlow, Keras и SciKit-Learn.

Исследование производилось на валидированных данных КТ 54 пациентов, предоставленных Национальной академией наук Беларуси. Они включали в себя снимки, содержащие патологии пневмонии и таких форм туберкулёза как диссеминированная, инфильтративная, очаговая, фиброзно-кавернозная и туберкулёма. С помощью программы VGG Image Annotator и медицинских специалистов Сибирского государственного медицинского университета было размечено и верифицировано 403 среза компьютерной томографии, использованных для обучения и тестирования моделей сегментации внутрилёгочных структур и классификации областей патологий.

Результатом проведённой работы стало предложенное и реализованное решение задачи локализации патологий диссеминированного туберкулёза лёгких на данных КТ, включающее в себя методы сегментации областей лёгочной ткани, регионов интереса и классификации патологий по текстурным признакам, обеспечивающее качество со значениями метрик: 0,61 – точность, 0,76 – полнота, 0,67 – F-мера.