АННОТАЦИЯ

на диссертационную работу по теме: «Разработка информационной модели и алгоритмов распознавания и машинного обучения рукописных текстов на казахском языке», докторанта PhD по специальности «6D075100 – Информатика, вычислительная техника и управление»
Елеусинова Армана Инабатовича

Актуальность работы: Распознавание рукописного текста (HTR) является ключом ко многим приложениям, которые зависят от оцифровки документов, включая обработку архивных документов, библиотечное дело, здравоохранение, банковское дело и т.д. Самая большая проблема возникает из-за большого разнообразия стилей. Методы машинного обучения предлагают потенциальное решение этой проблемы. Эти методы легко масштабируются для миллионов документов, они быстрее транскрипции человека и обеспечивают воспроизводимые результаты. Многие программные приложения уже реализуют HTR, но они не подходят для казахского языка из-за особенности языка, и исследования в этой области все еще актуальны.

Архив Президента Республики Казахстан является хранителем уникальных документов по новейшей политической истории Казахстана. По состоянию на 1 января 2020 г. в Архиве насчитывается более 830 тыс. единиц хранения за 1918-2008 гг. из них 762 тыс. единиц хранения на бумажной основе, в том числе написанных рукописном варианте.

Оцифровка архивных документов – это технически сложный процесс, который зависит от многих факторов: наличия специального программно-аппаратного комплекса, определения рациональной стратегии работ, квалификации операторов сканирования и наличия необходимой нормативно-методической документации. В настоящее время вместо термина «оцифровка» иногда употребляют схожее понятие – сканирование.

Оцифровка дает возможность безопасного и более экономичного хранения, удобного поиска необходимой информации, а также оперативного доступа к архивным документам. Оцифровка архивных документов позволяет решить одновременно несколько задач:

- наведению порядка в фондовом учете;
- обеспечение сохранности подлинников путем постепенного вывода их из оборота использования;
- расширение и упрощение доступа к документам архивного фонда;

- предоставление доступа к электронному фонду;
- ускорение процесса предоставления государственных услуг;
- работа с научно-справочным аппаратом архива

Для решения данных проблем требуется комплексное исследование и внедрение последних технологических и исследовательских результатов, включающих распознавание рукописных документов, где потребуются большие наборы специально собранных данных.

Цель исследования: Разработка алгоритмов машинного обучения и архитектуры нейронной сети для распознавания архивных рукописей на казахском языке

Задачи исследования:

Для достижения поставленной цели рассматриваются следующие задачи:

- Сбор и обработка базы данных для обучения и тестирования методов машинного обучения распознавания рукописного текста на казахском языке.
- Исследование и анализ систем, методов и алгоритмов распознавания рукописного текста;
- Разработка алгоритмов и моделей сегментации для предварительной обработки архивных документов;
- Разработка модели машинного обучения на основе генеративно-состязательных нейронных сетей (GAN) для искусственного дополнения данных рукописного текста на казахском языке;
- Разработка модели машинного обучения на основе трансформатора для распознавания текста;
- Разработка системы поиска информации в документах

Объект исследования: Изображения рукописных текстов на казахском языке.

Предмет исследования: Алгоритмы машинного обучения, нейронные сети и информационные модели для распознавания рукописных текстов на казахском языке

Методы исследования: Алгоритмы компьютерного зрения, глубокие нейронные сети, сверточные сети, генеративно-состязательные сети (GAN), архитектура трансформатора

Научная новизна работы:

- Разработана гибридная архитектура CNN+Transformer, адаптированная для распознавания рукописного текста на казахском языке.

- Предлагается метод, основанный на архитектуре GAN, для дополнения данных рукописного текста на казахском языке путем создания реалистичных вариаций почерка. Этот метод показал, что точность распознавания можно повысить даже при небольшом объеме данных.

Теоретическая значимость исследования: Исследование вносит вклад в развитие методов машинного обучения и нейронных сетей, используемых для распознавания рукописного текста. Разработка и адаптация архитектур (например, CNN, Transformer) к специфике казахской письменности расширяет теоретическую базу знаний в области интеллектуальной обработки текстовой информации.

В отличие от широко изучаемых языков казахский язык имеет агглютинативный строй, сложную морфологию и алфавитные особенности. Исследование позволяет теоретически обосновать и предложить новые методы распознавания и обработки текстов на казахском языке.

Практическая значимость полученных результатов: Разработка автономной системы интеллектуальной обработки документов является перспективным направлением развития информационно-телекоммуникационных технологий в Республике Казахстан. Внедрение результатов данного проекта позволит в первую очередь повысить эффективность процесса оцифровки рукописных текстов. Соответственно, основной социально-экономический эффект от внедрения результатов работы связан с экономией времени и ручного труда на обработку документов различных видов и объемов.

Положения диссертации, выносимые на защиту:

- Гибридная архитектура сверточной нейронной сети и трансформатора для распознавания рукописного текста позволяет снизить показатели wer (частота ошибок в словах) и cer (частота ошибок в символах) на 10,44% и 2,85% соответственно.
- Метод искусственного дополнения данных рукописного текста на основе GAN позволяет повысить точность распознавания слов с 75,8% до 86,3%.

Объем и структура работы: Диссертация состоит из введения, трех разделов, заключения, списка литературы и приложений. Полный объем диссертации составляет: 84 страниц машинописного текста, включая 54 рисунка, 6 таблиц, 83 ссылки и 2 приложения.

Во Введении дано описание проблемы, краткий обзор предметной области. Обоснована актуальность диссертационной работы, проведен обзор литературных источников.

В первой главе рассматривается проблема распознавания рукописного текста и общие подходы к ее решению. Подробно описаны методы создания и обработки наборов данных для распознавания рукописного текста. Показаны способы улучшения качества изображений и повышения точности распознавания с использованием методов предварительной обработки (фильтрация, выравнивание гистограммы, удаление шума). Этот этап считается важным подготовительным шагом для повышения эффективности распознавания текста.

Во второй главе рассматриваются современные методы машинного обучения, используемые для распознавания рукописного текста. Сначала представлены методы сегментации строк и слов, нейронные сети, которые их реализуют, а также результаты оценки. Далее описываются модели на основе архитектур CNN и Transformer для классификации букв и распознавания слов. Кроме того, описывается использование метода GAN (Generative Adversarial Networks) для увеличения объема данных и его результаты.

В третей главе описываются архитектура, компоненты и техническая реализация разработанной системы распознавания рукописного текста. Компоненты системы оптимизированы с использованием Nvidia Triton Inference Server. Кроме того, рассматривается система автоматического поиска и извлечения необходимой информации из документов, а также возможности интеграции и развертывания готовой системы.

Уровень достоверности и результаты апробации: Обоснованность и достоверность исследования соответствуют предъявляемым требованиям Основные положения диссертации прошли апробацию и изложены в 6 печатных работах, из них:

- 2 статьи в международных научных журналах, индексируемых в Scopus с высоким импакт-фактором Q2, Q3:

1. **Yeleussinov, A.**, Amirgaliyev, Y., Cherikbayeva, L. Improving OCR Accuracy for Kazakh Handwriting Recognition Using GAN Models. *Appl. Sci.* **2023**, *13*, 5677 (SJR 0.492, процентиль 75) https://doi.org/10.3390/app13095677

2. Amirgaliyev B., **Yeleussinov A.**, Taizo M. Kazakh handwritten recognition. Journal of Theoretical and Applied Information Technologies, 2020, 98(14), pp. 2744–2754.

- 1 статья в журнале из списка, рекомендованного Комитетом по обеспечению качества в сфере науки и высшего образования Министерства науки и высшего образования Республики Казахстан (КОКСНВО):

1. Амиргалиев Е. Н., Миачи Т., **Елеусинов А.И.** Методы и алгоритмы формирования базы и обработки рукописных текстов на казахском языке, Вестник КазНИТУ, 2020г., №138, сс. 653-660.

- 3 публикаций в материалах международных и республиканских конференций и докладов на научных семинарах:

1. **Yeleussinov A.**, Islamgozhayev T., Kozhirbayev Z. Creation and Annotation of a Handwritten Text Database for the Kazakh Language: Methodologies and Preliminary Results. In2024 International Conference on Electrical, Communication and Computer Engineering (ICECCE) 2024 Oct 30 (pp. 1-6). IEEE.)

2. **Елеусинов А.И.**, Бурибаев Ж.А. Сегментация рукописных символов с исправлением наклона текста. V-международная конференция «Информатика и прикладная математика», Алматы, ИИВТ, 29.09.2020-01.10.2020.

3. **Елеусинов А.И.,** Амиргалиев Е. Н., Мажитов Ш. С. О разработке базы изображений рукописных символов казахского языка, Конференция"IV международная научно-практическая конференция "Информатика и прикладная математика", посвященной 70-летнему юбилею профессоров Биярова Т.Н., Вальдемара Вуйцика и 60-летию профессора Амиргалиева Е.Н., Институт информационных и вычислительных технологий МОН РК, 25.09.2019-29.09.2019.

Диссертация обсуждалась также на научных семинарах кафедры.