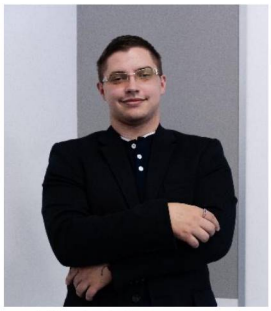
УДК 004.043

# ИССЛЕДОВАНИЕ ВЛИЯНИЯ ОРИЕНТАЦИИ ДОКУМЕНТА НА ТОЧНОСТЬ ИЗВЛЕЧЕНИЯ ДАННЫХ МУЛЬТИМОДАЛЬНОЙ МОДЕЛЬЮ PADDLEOCR-VL

***Е.А. Курлюк***
*Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР*
*kurluke750@gmail.com*



***Н.А. Ларченко***
*Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР*
*nikitadeve@gmail.com*

***А.В. Бойдич***
*Студент инженерно-экономического факультета кафедры экономической информатики БГУИР*
*andrewwboidich@gmail.com*

***М.В. Давыдов***
*Первый проректор, доцент, кандидат технических наук, БГИУР*
*davydov-mv@bsuir.by*

***Е.А. Курлюк***
*Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, оптимизацией извлечения и структурирования информации из медицинских документов, разработкой масштабируемых информационно-компьютерных систем и сервисов.*

***Н.А. Ларченко***
*Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, построением нейросетевых систем анализа физиологических сигналов, разработкой масштабируемых информационно-компьютерных систем и сервисов.*

***А.В. Бойдич***
*Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, построением и исследованием нейросетевых систем.*

***М.В. Давыдов***
*Первый проректор Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, цифровой обработкой сигналов, электростимуляцией и транскраниальной магнитной стимуляцией.*

**Аннотация.** В работе исследуется влияние ориентации изображения на качество извлечения табличной структуры мультимодальной моделью PaddleOCR-VL. Эксперименты проведены на подмножестве набора данных SynthTabNet с четырьмя фиксированными углами поворота. Показано, что изменение ориентации приводит к существенному снижению качества распознавания, особенно при повороте на 180°. Предложен этап предварительного определения угла поворота на основе классификатора ResNet34. Добавление данного этапа приводит к выравниванию распределений метрик TEDS и TEDS-struct и снижению разброса результатов. Полученные результаты демонстрируют эффективность включения отдельного модуля определения ориентации в конвейер обработки документов.

**Ключевые слова:** оптическое распознавание текста, мультимодальные модели, анализ структуры таблиц.

**Введение.** Оптическое распознавание текста (OCR) представляет собой совокупность методов и алгоритмов, направленных на извлечение текстовой информации из изображений и сканированных документов. OCR-системы широко применяются в задачах цифровизации архивов, автоматизации документооборота, обработки медицинских и финансовых записей, а также в системах интеллектуального анализа данных [1-4].

Точность распознавания зависит от качества входных изображений. На результат влияют шум, низкое разрешение, сложный фон и геометрические искажения. Существенным фактором является ориентация документа [5].

На практике изображения документов могут быть повернуты относительно исходной ориентации текста. Это возникает при сканировании, фотографировании документов без фиксации положения камеры, обработке потоков изображений. В таких случаях файл сохраняется без изменения, но текст оказывается повернут относительно горизонтальной оси. Малые углы поворота документа (5-10°) корректируются алгоритмами выравнивания, основанными на анализе текстовых линий [6]. Повороты на 90°, 180° и 270° не относятся к задаче коррекции наклона и требуют отдельного анализа, так как изменяют ориентацию текста на уровне всего изображения и влияют на результат распознавания [7, 8].

Современные подходы к распознаванию текста основаны на мультимодальных моделях, объединяющих визуальные признаки изображения и языковые зависимости текста [9, 10]. Такие модели учитывают локальные графические характеристики символов, контекст на уровне слов, строк и структуры документа. Это позволяет повысить точность извлечения данных в условиях шума, сложного фона и вариативности оформления документов. При этом влияние значительных поворотов изображения на качество работы подобных моделей остаётся недостаточно изученным. В данной работе рассматривается мультимодальная модель PaddleOCR-VL, разработанная в рамках экосистемы PaddleOCR [11]. Модель ориентирована на извлечение структурированной информации из документов и объединяет методы компьютерного зрения и обработки естественного языка. Несмотря на наличие в экосистеме PaddleX инструментов для коррекции ориентации изображений, в рамках настоящего исследования анализируется непосредственно поведение модели PaddleOCR-VL без использования дополнительных модулей выравнивания. В работе рассматривается модель PaddleOCR-VL, разработанная в рамках экосистемы PaddleOCR [11]. Модель предназначена для извлечения текстовой и структурированной информации из документов. В рамках экосистемы PaddleOCR модель PaddleOCR-VL извлекает структурированную текстовую информацию из документов посредством последовательного двухэтапного анализа. На первом этапе алгоритм генерирует визуальные признаки изображения и локализует текстовые области, на втором – выполняет посимвольное распознавание с формированием последовательностей. Итоговая обработка объединяет контекст текстовых данных с их пространственным расположением и визуальными признаками документа, что обеспечивает точное определение логической структуры и семантическую экстракцию данных.

Целью работы является исследование влияния ориентации документа на точность извлечения данных мультимодальной моделью PaddleOCR-VL.

**Методология проведения эксперимента.** Для исследования влияния ориентации изображения на качество разбора таблиц была выбрана двухэтапная схема эксперимента. На первом этапе оценивалась работа исходной модели без коррекции ориентации входных данных. На втором этапе перед подачей изображения в модель добавлялся отдельный модуль определения угла поворота, после чего выполнялась повторная оценка на том же наборе данных. Такой подход позволил раздельно оценить чувствительность модели к повороту изображения и эффект от предварительного определения и корректировки ориентации. Для проведения эксперимента использовалось готовое окружение в Docker с установленными зависимостями и предварительно загруженными весами PaddleOCR-VL. Все вычисления выполнялись на графическом ускорителе Nvidia RTX 4060.

При запуске PaddleOCR-VL использовалась фиксированная текстовая инструкция «Table Recognition:». Максимальное число генерируемых токенов задавалось равным 512. Для ускорения вычислений использовался Flash Attention 2, при этом расчёты выполнялись в формате FP16. Генерация проводилась в детерминированном режиме (стохастическая выборка была отключена), для каждого изображения формировалась одна выходная последовательность. Такая настройка исключала случайные флуктуации значений между запусками и позволяла получать сопоставимые результаты при сравнении разных углов поворота. Результат работы модели формировался в формате OTSL.

Оценка модели выполнялась на наборе данных SynthTabNet. Для анализа влияния ориентации использовался поднабор из 1000 изображений. Для каждого исходного изображения формировались изображения с четырьмя фиксированными ориентациями: 0°, 90°, 180° и 270°. Это позволяло оценить поведение модели при отсутствии поворота и при трёх дискретных вариантах изменения ориентации. На данном этапе коррекция положения изображения не выполнялась, каждое изображение в заданной ориентации напрямую подавалось в модель, после чего сохранялся результат распознавания таблицы.

Оценка качества выполнялась по метрике TEDS, определяемой как

$$TEDS=1-distancennodes, \quad (1)$$

где distance – расстояние редактирования между структурой предсказанной и эталонной таблиц, nnodes – число элементов в эталонной разметке. Также использовалась метрика TEDS-struct, оценивающая только структурное соответствие таблиц без учёта текстового содержимого.

Поскольку выход модели представлен в формате OTSL, было реализовано преобразование в HTML-представление таблицы. Преобразование включало распознавание структуры таблицы, определение ячеек, учёт объединений по строкам и столбцам и формирование корректной HTML-разметки. Полученная разметка использовалась для сравнения с эталонными данными. После завершения первого этапа был добавлен модуль определения угла поворота изображения. Его задача состояла в том, чтобы до запуска PaddleOCR-VL определить, к какому из четырёх классов относится ориентация документа: 0°, 90°, 180° или 270°. На основании этого предсказания изображение приводилось к правильному положению и затем повторно подавалось в PaddleOCR-VL с теми же параметрами, что и на первом этапе.

Для задачи обучения классификатора был сформирован набор данных для обучения модели определения угла поворота. В качестве исходного источника использовалась обучающая часть SynthTabNet. Для каждого изображения выполнялся детерминированный поворот в четыре класса: 0°, 90°, 180° и 270°. Общий объём составил 4000 изображений. Набор был разделён на обучающую, проверочную и тестовую части в соотношении 3200, 400 и 400 изображений соответственно. Для определения угла поворота была использована архитектура ResNet34 с инициализацией весами ImageNet. На вход модели подавалось изображение документа, на выходе предсказывался класс ориентации. Входное изображение приводилось к размеру 224×224. Обучение проводилось в течение 8 эпох с использованием оптимизатора AdamW, скоростью обучения 1e-4 и коэффициентом weight decay 1e-4. В качестве функции потерь использовалась кросс-энтропия. Таким образом, в обеих сериях эксперимента конфигурация основной модели оставалась неизменной, а различие между ними сводилось только к наличию или отсутствию предварительного приведения изображения к правильной ориентации.

**Обсуждение результатов.** Для оценки поведения модели при различных ориентациях изображения рассматриваются распределения значений метрик TEDS и TEDS-struct по каждому углу поворота. Такой способ представления позволяет сопоставить и средние значения, и характер распределения результатов для разной ориентации изображения. Распределения значений TEDS для PaddleOCR-VL без предварительного определения ориентации приведены на рисунке 1. Для изображений без поворота распределение

сосредоточено в области высоких значений, что соответствует устойчивому распознаванию структуры и содержимого таблиц. При поворотах на 90° и 270° распределения смещаются влево: увеличивается доля результатов со средними значениями метрики, а число примеров с высоким качеством уменьшается. Наиболее выраженное изменение наблюдается при повороте на 180°. В этом случае распределение теряет выраженную концентрацию в правой части шкалы, возрастает доля низких и околонулевых значений, а разброс становится существенно шире.

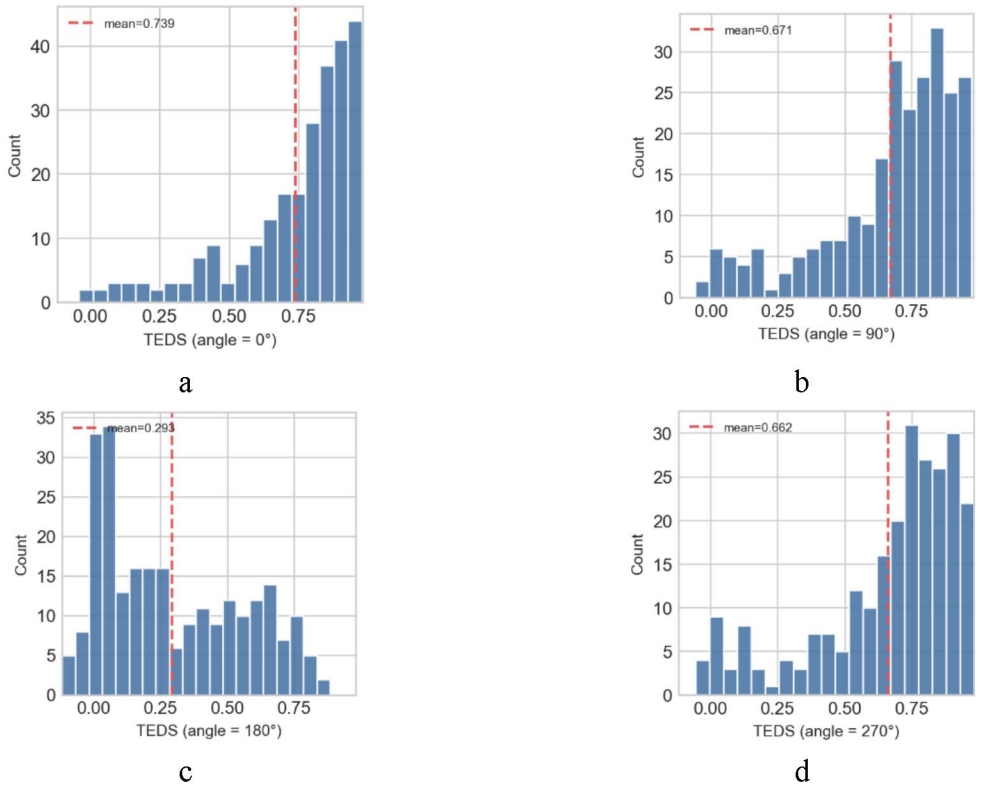



*Рисунок 1.* Распределение значений TEDS для PaddleOCR-VL без предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

Распределения значений TEDS-struct без предварительного определения ориентации приведены на рисунке 2. Для изображений с ориентацией 0° значения также сосредоточены в правой части шкалы. При поворотах на 90° и 270° сохраняется смещение в сторону меньших значений, однако изменение выражено слабее, чем для полной метрики TEDS. Для поворота на 180° снижение качества также фиксируется, но структура распределения остаётся более устойчивой, чем в случае TEDS. Это означает, что при перевороте изображения модель чаще сохраняет общий каркас таблицы, однако некорректно восстанавливает полное представление таблицы вместе с содержимым и внутренними связями между ячейками. Таким образом, влияние угла поворота сильнее проявляется на уровне полного структурно-текстового соответствия, чем на уровне одной только структуры. Для определения угла поворота обучен классификатор на основе ResNet34. Кривые обучения приведены на рисунке 3. Функция потерь на обучающей и валидационной выборках быстро снижается в первые эпохи и далее практически не изменяется. Значения точности и macro-F1 на валидационной выборке достигают максимальных значений на раннем этапе обучения и остаются стабильными. Точность на валидационной и тестовой выборках составила 1. При применении классификатора к набору данных, использованному в основном эксперименте, точность составила 0.997.

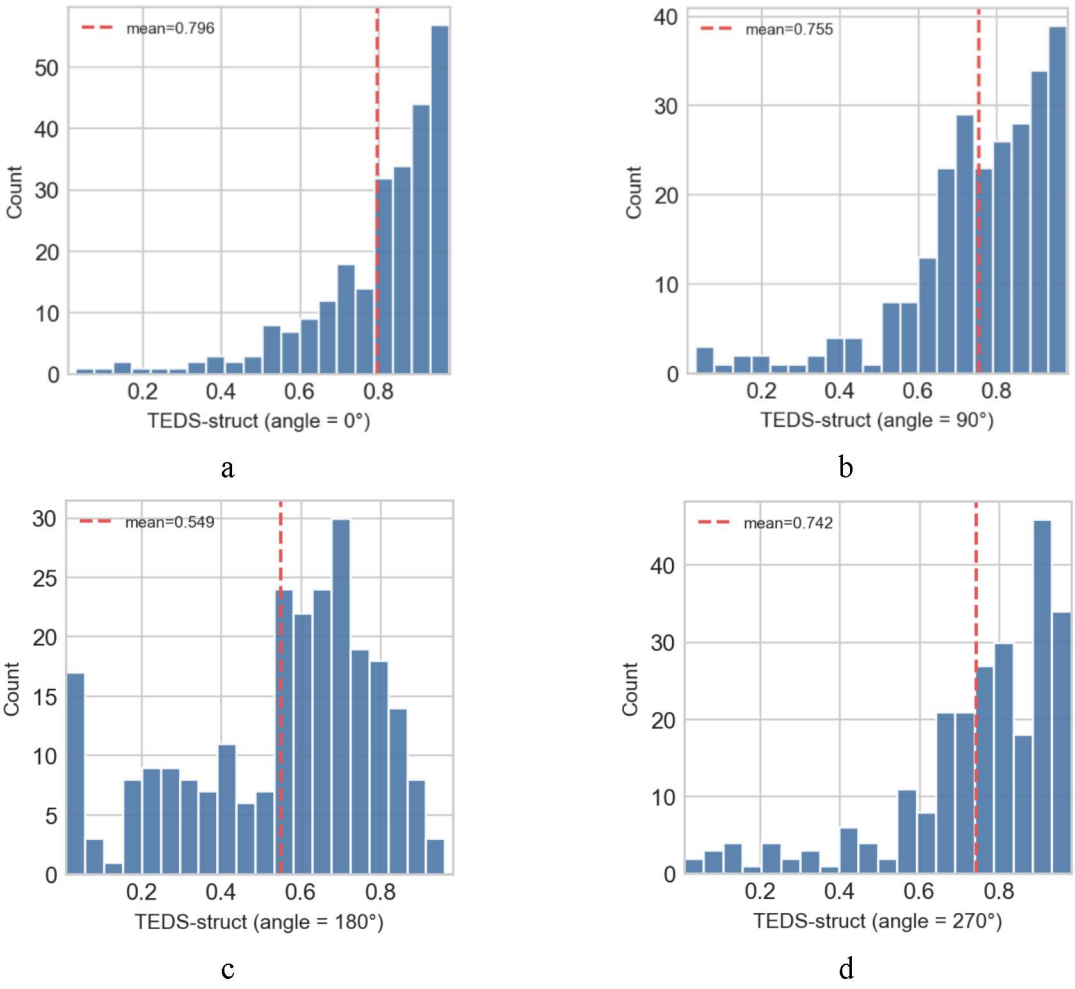



*Рисунок 2.* Распределение значений TEDS-struct для PaddleOCR-VL без предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

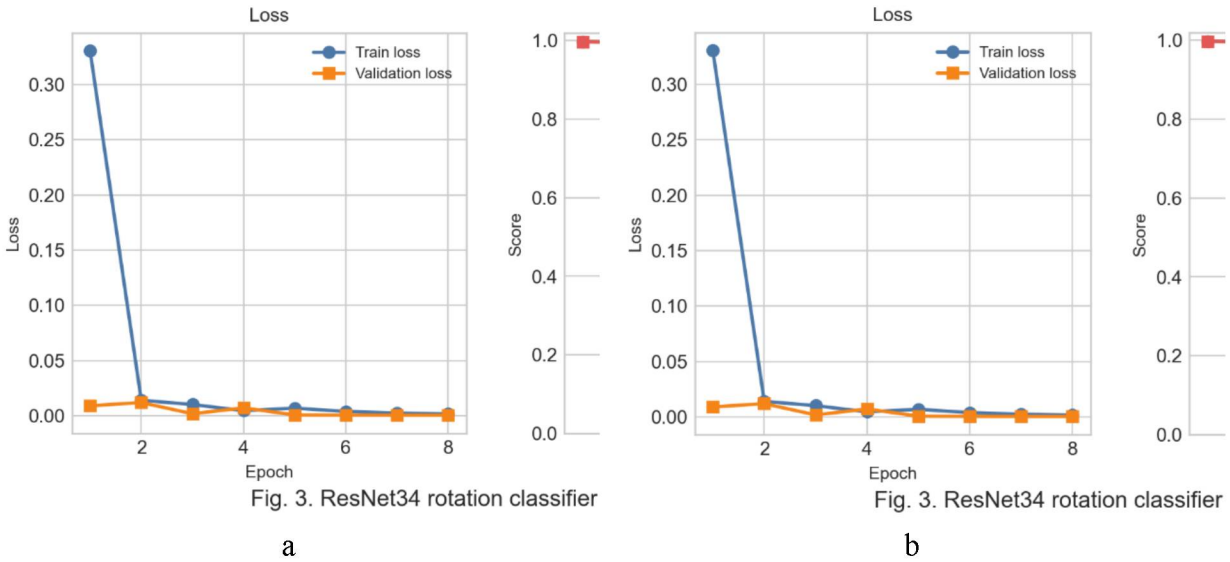



*Рисунок 3.* Кривые обучения классификатора ResNet34: а) функция потерь на обучающей и валидационной выборках; б) точность и macro-F1 на валидационной выборке.

После добавления классификатора ориентации перед запуском PaddleOCR-VL распределения TEDS изменяются существенным образом. Результаты приведены на рисунке 4.

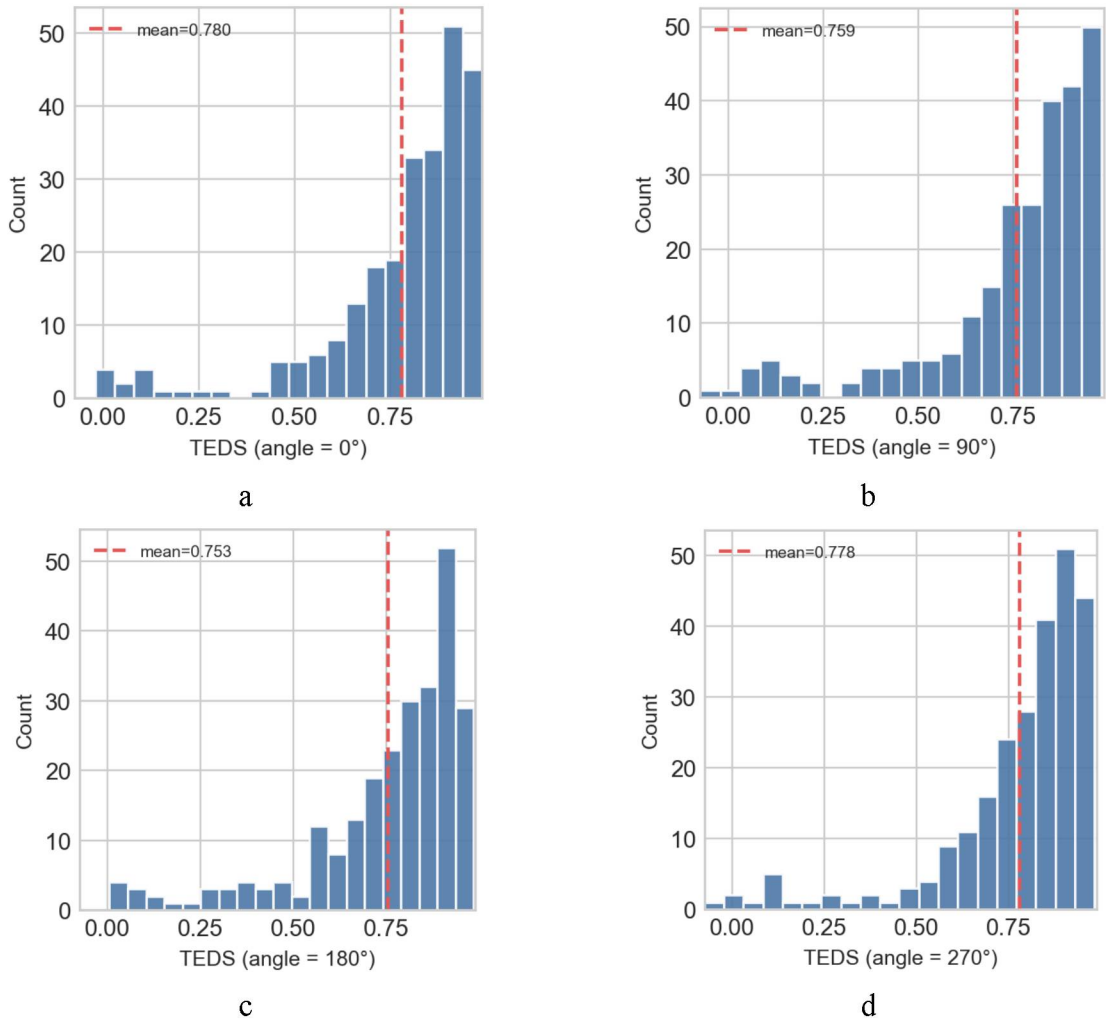



*Рисунок 4.* Распределение значений TEDS для PaddleOCR-VL после предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

Для всех четырёх углов распределения сосредоточены в области высоких значений и имеют близкую форму. Различие между поднаборами по углам поворота становится значительно менее выраженным. Наибольшее изменение относится к изображениям, повернутым на 180°: поднабор, ранее характеризовавшийся широким разбросом и заметной долей низких значений, после исправления ориентации демонстрирует распределение, сопоставимое с остальными углами. Это указывает на то, что существенная часть ошибок на первом этапе была связана не со сложностью таблиц как таковой, а с неправильным положением изображения относительно ожидаемой ориентации текста. Сопоставление результатов первого и второго этапов показывает, что предварительное определение угла поворота влияет не только на средние значения метрик, но и на форму распределений. На первом этапе качество распознавания зависело от угла поворота, причём наиболее выраженное снижение наблюдалось для изображений, перевёрнутых на 180°. После добавления классификатора ориентации распределения для всех четырёх углов становятся близкими как по положению, так и по форме. Это означает, что влияние ориентации изображения на итоговое качество распознавания в значительной степени снимается ещё

до запуска PaddleOCR-VL. В результате модель работает на более однородном входном наборе, а разброс качества между различными углами поворота существенно сокращается.

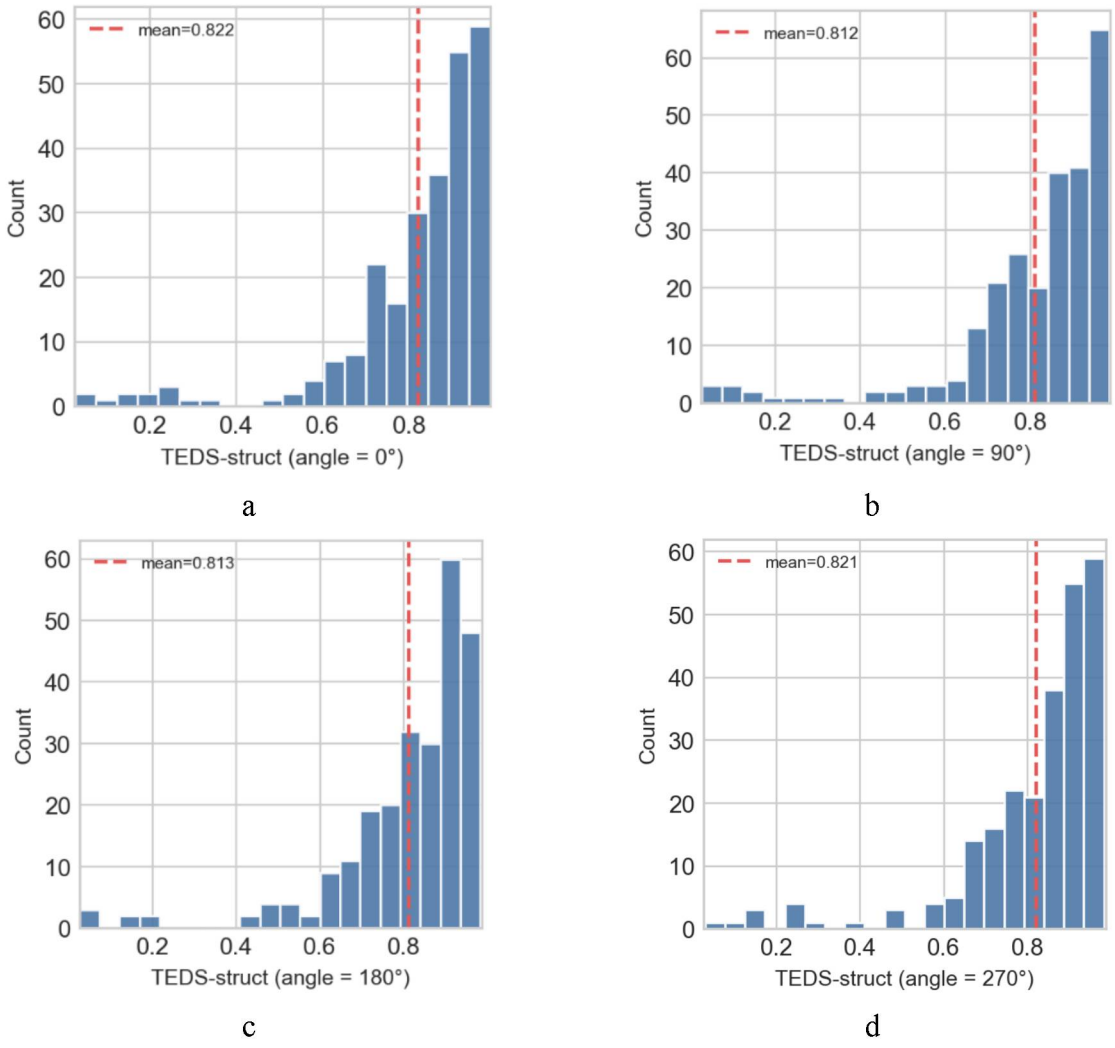



*Рисунок 5.* Распределение значений TEDS-struct для PaddleOCR-VL после предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

**Выводы.** Полученные результаты подтверждают, что ориентация изображения является значимым фактором, влияющим на качество извлечения табличной структуры. Изменение положения документа приводит к перераспределению значений метрик и увеличению доли некорректных предсказаний, что указывает на зависимость модели от исходной ориентации входных данных.

Введение этапа определения угла поворота изменяет характер работы системы. После приведения изображений к единой ориентации значения метрик стабилизируются, а различия между поднаборами, сформированными по углам поворота, практически исчезают. Это указывает на то, что часть ошибок обусловлена не сложностью самих таблиц, а некорректным положением текста на изображении относительно ожидаемой структуры. Отдельно следует отметить, что классификация ориентации решается с высокой точностью и не требует сложной модели или длительного обучения. Это делает возможным включение данного этапа в общий конвейер без существенного увеличения вычислительных затрат. Полученные результаты показывают, что использование специализированного предварительного этапа обработки позволяет повысить устойчивость мультимодальной модели без изменения её архитектуры.

**Список литературы**

[1] Fleischhacker, D.; Kern, R.; Göderle, W. Enhancing OCR in historical documents with complex layouts through machine learning // International Journal on Digital Libraries. – 2025. – Vol. 26.

[2] Liu, Y.; Li, Z.; Huang, M.; Zheng, Y.; Zhang, X.; Chen, L. OCRBench: on the hidden mystery of OCR in large multimodal models // Science China Information Sciences. – 2024. – Vol. 67. – Art. 220102.

[3] Ullah, H.; Tanveer, M.; Jan, A. Enhancing handwritten prescription recognition with AI-driven OCR // Journal of Computing & Biomedical Informatics. – 2025.

[4] Nagalaxmi, B.; Sujatha, P.; Sandeep, G.; Lakshmi Varun, G.; Harshavardhan Teja, B. Optimized OCR approaches for accurate text extraction in legal and financial document automation // Journal of Science Engineering Technology and Management Science. – 2025. – Vol. 2, No. 7. – P. 486–495.

[5] Курлюк, Е. А. Применение компьютерного зрения для автоматизированной обработки медицинских документов / Е. А. Курлюк, Н. А. Ларченко, М. В. Давыдов, Е. К. Курлянская // Цифровая трансформация. – 2025. – Т. 31, № 4. – С. 55–64.

[6] Soumya, B. J.; Vasudev, T. Enhancing document image processing: correcting skew in printed documents using deep learning // Journal of Information Systems Engineering and Management. – 2025. – Vol. 10, No. 25s.

[7] Tuggener, L.; Stadelmann, T.; Schmidhuber, J. Efficient rotation invariance in deep neural networks through artificial mental rotation // Frontiers in Computer Science. – 2025. – Vol. 7. – Art. 1644044.

[8] Goswami, S.; Ravi, A.; Kolla, R.; Faraz, A.; et al. Seeing straight: document orientation detection for efficient OCR // arXiv. – 2025. – DOI: 10.48550/arXiv.2511.04161.

[9] Ding, Y.; Han, S. C.; Lee, J.; et al. Deep learning based visually rich document content understanding: a survey // Artificial Intelligence Review. – 2026. – Vol. 59. – Art. 114.

[10] Ke, W.; Zheng, Y.; Li, Y.; Xu, H.; et al. Large language models in document intelligence: a comprehensive survey, recent advances, challenges, and future trends // ACM Transactions on Information Systems. – 2026. – Vol. 44, No. 1. – Art. 18.

[11] Cui, C.; Sun, T.; Liang, S.; Gao, T.; et al. PaddleOCR-VL: Boosting multilingual document parsing via a 0.9B ultra-compact vision-language model // arXiv. – 2025. – DOI: 10.48550/arXiv.2510.14528.

**Авторский вклад**

**Курлюк Евгений Александрович** – постановка задачи исследования, разработка методологии оценки эффективности предварительной обработки изображений при извлечении текстовой информации; проектирование и обучение классификатора на базе ResNet34; анализ влияния геометрических искажений на метрики TEDS и TEDS-struct.

**Ларченко Никита Александрович** – постановка задачи исследования, подготовка выборок данных на основе датасета SynthTabNet; проведение испытаний PaddleOCR-VL на всех этапах исследования, участие в сборе статистических данных, формулировка выводов и направлений для дальнейших исследований.

**Бойдич Андрей Викторович** – постановка задачи исследования; сравнительный анализ распределений метрик до и после внедрения модуля коррекции; интерпретация результатов распознавания табличных структур и подготовка графических материалов исследования.

**Давыдов Максим Викторович** – постановка задачи исследования; руководство исследованием по оценке устойчивости мультимодальных моделей к изменению ориентации входных данных; формулировка выводов и направлений для дальнейших исследований.

# STUDY OF THE IMPACT OF DOCUMENT ORIENTATION ON DATA EXTRACTION ACCURACY USING THE PADDLEOCR-VL MULTIMODAL MODEL

***E.A. Kurlyuk***
*Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR*

***N.A. Larchenko***
*Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR*

***A.V. Boidzich***
*Student of the Faculty of Engineering and Economics, Department of Economic Informatics BSUIR*

***M.V. Davydov***
*First Vice-Rector, PhD, BSUIR*

**Abstract.** This paper investigates the impact of image orientation on table structure extraction quality using the PaddleOCR-VL multimodal model. Experiments are conducted on a subset of the SynthTabNet dataset with four fixed rotation angles. The results show that changes in orientation lead to a significant degradation in recognition quality, particularly for 180° rotations. A preprocessing stage based on a ResNet34 rotation classifier is introduced to determine the image orientation prior to inference. Incorporating this stage results in more consistent TEDS and TEDS-struct score distributions and reduces performance variability. The findings demonstrate the effectiveness of integrating a dedicated orientation detection module into the document processing pipeline.

**Keywords**: OCR, multimodal models, table structure analysis.