Классификация с использованием нейронных сетей изображений символов церковнославянского языка XI–XVII веков

А.М. Егорова [1], Д.В. Демидов [1], А.Д. Егоров [1]
[1] Национальный исследовательский ядерный университет «МИФИ»,115409, Россия, г. Москва, Каширское ш., д. 31

Аннотация

В работе рассматривается задача распознавания символов церковнославянского языка в рукописях XI–XVII веков с использованием методов машинного обучения. В рамках исследования была создана база данных, содержащая 20 180 изображений 48 различных букв, полученных из 647 разворотов богослужебных книг. Для устранения дисбаланса классов использовались методы аугментации данных.

Для классификации символов протестированы различные модели нейросетей, среди которых наилучший результат показала EfficientNet-B4 с метрикой F1-weighted 0,976911. С учётом вычислительной эффективности и времени обработки был рассчитан интегральный критерий качества, согласно которому наиболее подходящей моделью для работы признана EfficientNet-B1. Данная модель обеспечивает скорость обработки 43 буквы в секунду, что соответствует 20 секундам на один разворот текста.

Результаты исследования демонстрируют возможность эффективного автоматизированного распознавания церковнославянских текстов, что значительно ускоряет их анализ и обработку. В дальнейшем планируется расширение базы данных и улучшение методов предобработки изображений для повышения точности распознавания.

Ключевые слова: искусственный интеллект, классификация изображений символов.

Цитирование: **Егорова, А.М.** Классификация с использованием нейронных сетей изображений символов церковнославянского языка XI–XVII веков / А.М. Егорова, Д.В. Демидов, А.Д. Егоров // Компьютерная оптика. – 2026. – Т. 50, № 4. – С. 1751. – doi:10.18287/COJ1751.

Citation: Egorova AM, Demidov DV, Egorov AD. Classification of Church Slavonic Letter Images from the 11th–17th Centuries Using Neural Networks. Computer Optics 2026; 50(4): 1751. doi:10.18287/COJ1751.

Введение

Распознавание текста (оптическое распознавание образов, OCR) является одной из важных задач, решаемых с помощью технологий машинного обучения. Технологии распознавания текстов на современных языках не просто достигли высокой точности распознавания текстов, но и позволяют распознавать различные тексты на естественном языке с точностью до 99 % [1] в зависимости от условий. Однако тематика распознавания текстов, написанных на архаичных языках, менее изучена. Хотя в ряде работ для некоторых языков удаётся добиться точности распознавания более 99 % [2, 3], тем не менее средняя точность распознавания древних рукописей в современных работах (2017–2023 годы) составляет порядка 93 % [4, 5]. При этом достичь большой точности мешает ряд фундаментальных преград [6], таких как большая зашумлённость и низкое качество страниц текста, обусловленные физическими дефектами, ветхостью и старостью материалов; качеством написания символов; сложностью подготовки соответствующей размеченной базы данных.

Глубина разработанности темы по языкам неоднородна. Так, существует всего несколько работ, посвящённых распознаванию старорусских, славянских, церковнославянских символов [3, 7, 8, 9]. Тем не менее в древних текстах можно найти много важной информации для филологов и историков, позволяющей первым исследовать развитие языка и текстов в динамике, а последним точнее датировать рукописи и определять суть происходящих событий. Эти задачи решаются сейчас и вручную, но автоматизированное распознавание может придать этой области исследований мощнейшее ускорение за счёт создания корпусов с инструментарием полнотекстового поиска [10].

Для решения любой задачи машинного обучения необходимо иметь собранную и специальным образом размеченную базу изображений. Такие базы в открытом доступе отсутствуют, несмотря на наличие в свободном доступе различных оцифрованных рукописей [11, 12].

При рассмотрении задач распознавания текстов на архаичных языках обычно не ставятся требования по производительности. Тем не менее в работе [13] указано, что среднее время обработки сканированной копии одной страницы текста составляет 11 секунд. Рассмотрение вопроса быстродействия подобных систем становится необходимым в связи с тем, что, с одной стороны, у исследователей отсутствует доступ к вычислительным мощностям, сопоставимым по объёму с теми, что используются в ИТ-индустрии для решения задач распознавания, при том что, с другой стороны, современные модели требуют всё больше вычислительных ресурсов, а с третьей стороны, растёт число оцифрованных текстов, при работе с которыми нужно решать различные задачи распознавания.

В данной работе решается задача высокопроизводительной классификации символов церковнославянского языка в рукописях XI–XVII веков. Для решения этой задачи подготовлена база изображений, а также проведено тестирование различных моделей машинного обучения.

1. Анализ существующих работ по распознаванию древних текстов

Сама по себе задача распознавания текста подразумевает решение двух подзадач: сегментации текста (разделения на буквы) и последующей классификации. В данной работе рассматривается только этап классификации.

В работе [14] показано, что в большинстве современных работ, посвящённых распознаванию текста на древних языках, используются различные нейросетевые модели. Для решения различных задач в основном применяются свёрточные, рекуррентные нейронные сети и модели-трансформеры.

В работе [15] для распознавания различных рукописных древних языков, например, языка Урду, используется свёрточная нейронная сеть на основе готовой модели VGG-16 [16]. С помощью модели удалось добиться точности в 98,66 % при значении F1-score 0,9864. При этом распознавалось всего 10 символов при наличии 8020 чётких незашумлённых изображений. В работе [17] также используется свёрточная нейросетевая модель, но разработанная самими авторами. Объектами классификации являются символы языка деванагари. Собранная авторами база в 9845 изображений была подготовлена на основе 400 различных текстов с помощью предварительной автоматизированной сегментации. Итоговая точность работы модели составила 94,6 %. В работе [14] также показано, что большая часть работ по анализу древних языков посвящена анализу текстов на латыни и древнегреческом языке.

Отдельно стоит рассмотреть работы, посвящённые церковнославянскому языку. В работе [3] также используется свёрточная нейросетевая модель для классификации букв. При этом все буквы в имеющейся базе данных имеют чёткую позицию на изображении, не зашумлены. Итоговая точность работы модели составила 99,13 %.

В работах [7, 18] проводится сравнение «классических» моделей – нечёткой классификации и классификации на основе дерева решений. Точность составила 78 и 76 процентов соответственно. При этом стоит отметить, что относительно низкая точность обусловлена не самими методами машинного обучения, а технологией извлечения признаков, которая сводилась к простой передаче пикселей изображения в модель для классификации, тогда как свёрточные нейронные сети проводят последовательную обработку признаков и сокращение пространства признаков. В работе [8] рассмотрена классификация старославянских букв боснийского и македонского алфавитов. В работе использовались деревья решений и нечёткая логика, а итоговая точность составила 82 %. В работе [19] для классификации церковнославянских букв также используется нечёткая логика. Итоговая точность составила 85 %, однако для классификации использовалось всего 16 букв. В работе [9] используются нейросетевые модели для достижения точности классификации в 96 %. При этом в качестве объектов классификации выступают символы в текстах XI–XVII веков. База данных в открытом доступе отсутствует.

Важно отметить, что во всех рассмотренных работах для предобработки изображений букв используется технология бинаризации изображений, которая оказывается применима в силу однородности рассмотренных источников (качественный фон, чёткие буквы) текстов на древних языках.

Таким образом, можно заключить, что характерным и наиболее простым методом распознавания древних языков является применение свёрточных нейронных сетей. Такие сети позволяют добиться средней точности в 93 %, а максимальной точности в 99 %. При этом значительную роль в работе моделей играют правильно подобранные данные.

2. Модели, методы и информационное обеспечение классификации символов древнего языка

Общая логика решения задачи распознавания текста требует решить две подзадачи: составить правильный набор данных и обучить наиболее подходящую модель машинного обучения, которая будет выбрана из числа различных моделей свёрточных нейронных сетей.

2.1. Подготовка базы изображений

На портале Корпуса рукописного наследия Древней Руси: https://www.slavcorpora.ru/ [20], разрабатываемом в НИЯУ МИФИ, создаётся база церковнославянских текстов XI–XVII веков. На данный момент в ней присутствует порядка 250 богослужебных книг – миней, доступных для зарегистрированных пользователей. Акцент на богослужебных книгах сделан в связи с тем, что в современных государственных и церковных библиотеках, архивах и фондах их сохранилось больше всего. 6 миней было сегментировано в полуавтоматическом режиме филологами Института русского языка им. В.В. Виноградова РАН с применением CVAT [10, 20]:

- Минея служебная, на ноябрь [рукопись]. – [Б. м.], конец XIV в. – 263 л. [из них 8 листов бумажных и 2 листа чистых], (25,9×19,7) см. Хранилище: Российская государственная библиотека.
- Минея служебная. Август. – Москва, 1497 г. – 163 л, (28×21) см. Хранилище: Российская государственная библиотека.

- Минея служебная. Март. – Новгород, 1369 г. – 133 л., (24,5×18) см. Хранилище: Российская национальная библиотека.
- Минея служебная. Июнь. – Новгород, 1439 г. – 236 л., (24,5×16,5) см. Хранилище: Российская национальная библиотека.
- Минея служебная, ноябрь-декабрь. Измайло, при псковском князе Феодоре Патрикееве. 1425 г. Хранилище: Государственный исторический музей.
- Минея служебная. Сентябрь. – 1095 г. см. Хранилище: Российский Государственный архив древних актов.

На рис. 1 приведён пример одного из использованных разворотов. Всего в работе было задействовано 647 разворотов из шести различных миней. Из каждого разворота по сегментированным областям были сформированы отдельные изображения под каждый символ. Примеры изображений букв показаны на рис. 2. В результате сегментации было получено 20 180 изображений, относящихся к 48 различным буквам алфавита (полный перечень классов приведён в Приложении В, рис. 5). Распределение количества изображений по буквам оказалось неравномерным: минимальное число экземпляров на класс составило 10, максимальное – 1270, медиана – 350. Подробная статистика по классам представлена в табл. 3 Приложения С.

Рис. 1. Пример разворота минеи

2.2. Аугментация данных

В связи с тем, что буквы используются в языке неравномерно, число изображений различных букв также оказывается различным. Для устранения дисбаланса классов различных букв используются методы аугментации данных [21]. Суть аугментации сводится к созданию из исходных образцов данных дополнительных синтетических образцов, которые могли бы присутствовать среди реальных примеров. Наличие синтетических данных позволяет повысить устойчивость модели к различным изменениям в данных (шум, искажение перспективы, нечеткость текста и т.д.).

В работе были использованы следующие методы аугментации:

1. Поворот: позволяет поворачивать изображения в пределах ± 30 градусов. Данная аугментация позволяет имитировать искажения при сканировании рукописи, связанные с наклоном страницы относительно вертикальной оси.
2. Смещение: случайное смещение изображения по горизонтали или вертикали на расстояние до 20 % от ширины изображения. Это позволяет имитировать сдвиг буквы при распознавании относительно геометрического центра изображения.
3. Деформация: случайное изменение пропорций изображения, а также положения углов изображения относительно друг друга. Это позволяет имитировать искажения при сканировании, связанные с изменением визуальной перспективы изображения.
4. Масштабирование: увеличение или уменьшение изображения на ± 20 %. Это позволяет имитировать различие размеров букв, что распространено в реальных рукописях.
5. Цветокоррекция: случайно изменяет цветовые параметры изображения (яркость, контраст и насыщенность). Данная аугментация имитирует различные варианты старения краски и носителя, на который нанесены буквы, а также разные варианты использованных чернил.
6. Размытие: гауссово размытие изображения. Данная аугментация имитирует нечёткое сканирование страниц.

В ходе работы было принято решение провести аугментацию данных с целью увеличения количества изображений в классах с недостаточным числом примеров. В результате общее число изображений составило 21 945, при этом в каждом классе оказалось не менее 100 изображений. Подробное распределение количества изображений по классам до и после аугментации представлено в Приложении С.

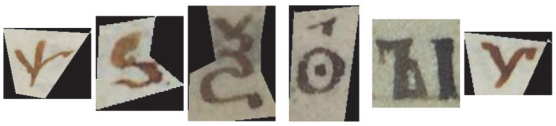

Рис. 2. Примеры изображений редких букв

2.3. Методика оценки качества классификации

Для оценки качества работы модели машинного обучения используется классическая F-мера (F1) [21]. В случае наличия в задаче многоклассовой классификации (каждая буква – отдельный класс) метрика должна быть адаптирована. Для этого в работе было использована **F1-weighted.** Средневзвешенная оценка F1 учитывает относительное число объектов в каждом из классов.

$$F1_{weighted} = \frac{\sum_{i=1}^{N}(F1_i * support_i)}{\sum_{i=1}^{N} support_i},$$

где N – количество классов, $support_i$ – количество изображений на класс i.

Значение F-меры принадлежит отрезку от 0 до 1. Чем ближе значение к единице, тем выше качество работы модели.

2.4. Сравниваемые модели классификации изображений

При решении задач компьютерного зрения для широкого использования доступно два различных типа нейросетевых моделей машинного обучения: трансформеры [23] и свёрточные нейронные сети [24]. В обзорах показано, что трансформеры демонстрируют [25] не меньшую точность в решении задачи распознавания изображений, чем свёрточные сети, однако их производительность ниже в силу большего числа весов и сложности архитектуры. В частности, трансформеры имеют квадратичную сложность вычислений в зависимости от размера входного изображения [26], тогда как сложность вычисления в свёрточных нейронных сетях растёт линейно в зависимости от размера изображений [27]. С учётом того, что свёрточные нейронные сети позволяют добиться почти максимальной точности, но при этом работают быстрее, модели-трансформеры в данной работе рассматриваться не будут.

Для сравнения были выбраны следующие архитектуры моделей: ResNet [28] с числом слоёв 18, 34, 50, 101, 152; EfficientNet [29] B0-B7; ResNeXt [30] 50 32×4d, 100 32×4d, 100 64×4d. В качестве оптимизатора был использован AdamW [31]. Каждая модель обучалась 15 эпох (эксперименты показали, что 15 эпох достаточно для качественной классификации). Для тестирования моделей используется технология отложенной выборки.

3. Анализ полученных результатов

Лучший результат среди сравниваемых моделей получен при использовании EfficientNet-B4, при её использовании достигается 0,976911 для F1-weighted. Результаты сравнения моделей представлены в табл. 1.

Табл. 1. Сравнение результатов работы различных моделей

Модель	F1-weighted на тестовой выборке
EfficientNet-B4	0,97691
EfficientNet-B1	0,9753
EfficientNet-B2	0,97486
ResNeXt-101 64×4d	0,97301
EfficientNet-B5	0,9726
EfficientNet-B3	0,97287
EfficientNet-B6	0,97173
EfficientNet-B7	0,97041
EfficientNet-B0	0,96903
ResNet-34	0,96582
ResNet-18	0,96482
ResNeXt-50 32×4d	0,96392
ResNeXt-101 32×8d	0,96197
ResNet-152	0,96144
ResNet-101	0,9591
ResNet-50	0,9588

На рис. 3 можно видеть соответствие времени обработки моделью одного изображения с буквой и качества работы модели. Видно, что в целом при близком качестве одни и те же модели могут работать с очень разной скоростью.

Для поиска наилучшего варианта построим интегральный критерий качества Q, который будет рассчитываться для каждой протестированной модели по следующей формуле:

$$Q^i = T^i_{norm} + M^i_{norm}, \text{где } T^i_{norm} = \frac{T^i}{T_{max}},$$

T^i – время обработки одного изображения (одного символа) i-й моделью, T_{max} – максимальное время обработки одного изображения (одного символа) среди всех протестированных моделей,

$$M^i_{norm} = \frac{M^i_{dif}}{M_{dif\,max}}, M^i_{dif} = \frac{1 - M^i}{\max(1 - M^i)},$$

M^i – значение $F1 - weighted$ для i-й модели.

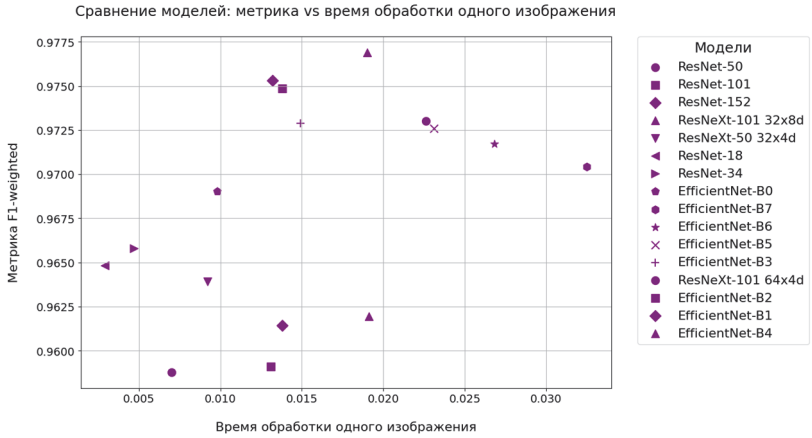

Рис. 3. График сравнения моделей по качеству и времени обработки одного изображения

Чем меньше значение критерия качества, тем эффективнее работает модель. В табл. 2 представлен расчёт критерия качества. Исходя из него, предпочтительной моделью является EfficientNet-B1. С её помощью получается обрабатывать 43 буквы в секунду, а общее время обработки одного разворота составит 20 секунд (один разворот в среднем содержит 860 букв). Подобная скорость обработки сопоставима с результатами в [13].

На рис. 4 представлен график зависимости метрики $F1_{weighted}$ от числа эпох обучения, на котором видно, что для обучения выбрано достаточное число эпох.

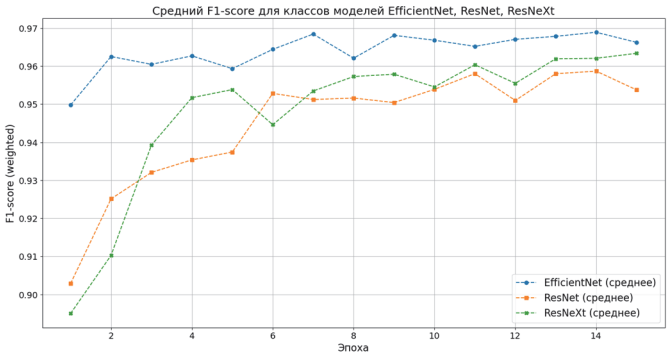

Рис. 4. График сравнения класса моделей по качеству

4. Обсуждение результатов

Современные нейросетевые модели позволяют добиться близкой к 100 % точности распознавания на изображениях церковнославянских букв, в том числе и начала прошлого тысячелетия. Ранее исследований с такими древними текстами не проводилось. При этом для успешного функционирования моделей не нужно использовать дополнительную обработку изображения типа бинаризации. Большинство работ не учитывает скорость обработки одной буквы в качестве значимого параметра работы модели, однако в реальности этот параметр оказывает большое влияние на практическую эффективность работы моделей. При наличии качественно отсканированных разворотов распознавание букв одного полного текста можно провести за полчаса–час.

Главная слабость всех проводимых исследований (включая и текущее) в том, что доступные тексты практически не содержат значимых искажений. Тексты с искажениями почти невозможно перевести в цифровой вид, не повредив источник. Таким образом, говорить о применении методов анализа с помощью машинного обучения в режиме реального времени не представляется возможным.

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации в рамках программы развития «Приоритет-2030 НИЯУ МИФИ».

References

[1] Chen X, Jin L, Zhu Y, Luo C, Wang T. Text recognition in the wild: A survey. ACM Computing Surveys 2021; 54(2): 1-35. DOI: 10.1145/3440756.
[2] Singh J, Lehal GS. Comparative performance analysis of feature(s)–classifier combination for Devanagari optical character recognition system. International Journal of Advanced Computer Science and Applications 2014; 5(6): 37-42. DOI: 10.14569/IJACSA.2014.050608.
[3] Smirnova AA, Grafeeva NG, Tokman MA. Recognition of Church Slavonic texts using machine learning methods. Pattern Recognition and Image Analysis 2024; 34(1): 212-218. DOI: 10.1134/S1054661824010209.
[4] Vijayalakshmi R, Gnanasekar JM. A review on character recognition and information retrieval from ancient inscriptions. In: 2022 8th International Conference on Smart Structures and Systems (ICSSS). IEEE; 2022: 1-7. DOI: 10.1109/ICSSS54384.2022.9788041.
[5] Krithiga R, Roja M, Deepa V, Sridevi T, Gowri P. Ancient character recognition: A comprehensive review. IEEE Access 2023. DOI: 10.1109/ACCESS.2023.3264102.
[6] Narang SR, Jindal MK, Kumar M. Ancient text recognition: A review. Artificial Intelligence Review 2020; 53(8): 5517-5558. DOI: 10.1007/s10462-020-09824-5.
[7] Martinovska C, Ivanovski Z, Veleski S, Cepreganova K. Recognition of Old Cyrillic Slavic letters: Decision tree versus fuzzy classifier experiments. In: 2012 6th IEEE International Conference on Intelligent Systems (IS). IEEE; 2012: 48-53. DOI: 10.1109/IS.2012.6335112.
[8] Martinovska Bande C, Veleski S, Cepreganova K, Ivanovski Z. Recognition features for Old Slavic letters: Macedonian versus Bosnian alphabet. International Journal of Scientific and Research Publications 2015; 5(12): 145-153. ISSN 2250-3153.
[9] Rabus A. Recognizing handwritten text in Slavic manuscripts: A neural-network approach using Transkribus. Scripta & e-Scripta 2019; 19: 9-32.
[10] Demidov DV, Lyapin IE, Boriskina EI, et al. Creating a corpus of handwritten heritage of Ancient Rus. In: Proceedings of the International Conference “Corpus Linguistics – 2023”. Saint Petersburg: Saint Petersburg State University Press; 2024: 72-81.
[11] Digitized collection of manuscripts and old printed books, GORAZD. Available at: http://gorazd.org/?q=ru/node/106 .
[12] Collections, National Library St. Kiril and Metodii. Available at: https://www.nationallibrary.bg/www/our-collections/ .
[13] Reul C, Springmann U, Puppe F. LAREX: A semi-automatic open-source tool for layout analysis and region extraction on early printed books. In: Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage. 2017: 137-142. DOI: 10.1145/3078081.3078094.
[14] Sommerschield T, Bodard G, Youn CM, Terras M, Mambrini F, McGillivray B. Machine learning for ancient languages: A survey. Computational Linguistics 2023; 49(3): 703-747. DOI: 10.1162/coli_a_00473.
[15] Chauhan VK, Singh S, Sharma A. HCR-Net: A deep learning based script independent handwritten character recognition network. Multimedia Tools and Applications 2024: 1-35. DOI: 10.1007/s11042-024-19249-x.
[16] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556; 2014.
[17] Moudgil A, Sharma S, Arora S, Dhir R. Handwritten Devanagari manuscript characters recognition using CapsNet. International Journal of Cognitive Computing in Engineering 2023; 4: 47-54. DOI: 10.1016/j.ijcce.2023.02.002.
[18] Klekovska M, Dimeski K, Najdenov B, Gievska S. Comparison of models for recognition of Old Slavic letters. In: ICT Innovations 2012: Secure and Intelligent Systems. Berlin, Heidelberg: Springer; 2013: 129-139. DOI: 10.1007/978-3-642-37169-1_12.
[19] Kuchuganov A, Lodygyn V, Malevich T, Zolotarev S. Automation of recognizing Old Slavonic manuscripts with help of description logics. Slavistica Vilnensis 2018; 63: 339-352.
[20] Corpus of Handwritten Heritage of Ancient Rus. Available at: https://www.slavcorpora.ru/ .
[21] Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In: Australasian Joint Conference on Artificial Intelligence. Berlin, Heidelberg: Springer; 2006: 1015-1021. DOI: 10.1007/11941439_114.
[22] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Advances in Neural Information Processing Systems 30 (NeurIPS 2017); 2017. arXiv:1706.03762.
[23] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998; 86(11): 2278-2324. DOI: 10.1109/5.726791.
[24] Jamil S, Jalil Piran M, Kwon OJ. A comprehensive survey of transformers for computer vision. Drones 2023; 7(5): 287. DOI: 10.3390/drones7050287.
[25] Keles FD, Wijewardena PM, Hegde C. On the computational complexity of self-attention. In: International Conference on Algorithmic Learning Theory (ALT 2023). Proceedings of Machine Learning Research; 2023: 597-619.
[26] Pan X, Zhan X, Dai B, Lin D, Loy CC. On the integration of self-attention and convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022: 815-825. DOI: 10.1109/CVPR52688.2022.00086.
[27] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016: 770-778. DOI: 10.1109/CVPR.2016.90.
[28] Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning (ICML). Proceedings of Machine Learning Research; 2019: 6105-6114.
[29] Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017: 1492-1500. DOI: 10.1109/CVPR.2017.634.
[30] Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101; 2017. DOI: 10.48550/arXiv.1711.05101.

Приложение А

Табл. 2. Расчет критерия качества исследуемых моделей

Модель	M^i	T^i	T^i_{norm}	M^i_{dif}	M^i_{norm}	Q
EfficientNet-B1	0,975302	0,0132	0,406153846	0,016505	0,911173678	0,494980168
EfficientNet-B2	0,974863	0,0138	0,424615385	0,016066	0,88693828	0,537677105
EfficientNet-B4	0,976911	0,019	0,584615385	0,018114	1	0,584615385
EfficientNet-B3	0,97287	0,0149	0,458461538	0,014073	0,776912885	0,681548653
EfficientNet-B0	0,969032	0,0098	0,301538462	0,010235	0,565032571	0,73650589
ResNet-18	0,964821	0,0029	0,089230769	0,006024	0,33256045	0,756670319
ResNet-34	0,965815	0,0047	0,144615385	0,007018	0,387435133	0,757180252
ResNeXt-101 64x4d	0,973005	0,0226	0,695384615	0,014208	0,784365684	0,911018931
EfficientNet-B5	0,9726	0,0231	0,710769231	0,013803	0,762007287	0,948761944
ResNeXt-50 32x4d	0,963918	0,0092	0,283076923	0,005121	0,282709506	1,000367417
EfficientNet-B6	0,971731	0,0268	0,824615385	0,012934	0,714033344	1,11058204
ResNet-50	0,958797	0,007	0,215384615	0	0	1,215384615
ResNet-152	0,96144	0,0138	0,424615385	0,002643	0,145909241	1,278706143
EfficientNet-B7	0,970414	0,0325	1	0,011617	0,64132715	1,35867285
ResNet-101	0,959101	0,0131	0,403076923	0,000304	0,016782599	1,386294324
ResNeXt-101 32x8d	0,961967	0,0191	0,587692308	0,00317	0,17500276	1,412689547

Приложение В

Рис. 5. Алфавит датасета

Приложение С

Табл. 3. Распределение символов в собранном датасете

№	Символ	Кол-во	№	Символ	Кол-во	№	Символ	Кол-во	№	Символ	Кол-во
1	а	977	13	и палочка	335	25	омега	253	37	х	358
2	б	440	14	иже	1270	26	от	492	38	ц	286
3	в	863	15	ижица	85	27	п	547	39	ч	342
4	г	505	16	йотированная а	523	28	пси	10	40	ш	257
5	д	573	17	йотированная е	260	29	р	537	41	щ	267
6	е	520	18	к	428	30	с	1051	42	ъ	522
7	е как эпсилон	167	19	кси	12	31	т	715	43	ы с ъ	63
8	е широкое	234	20	л	412	32	у	282	44	ы с ь	185
9	ж	406	21	м	750	33	ук	133	45	ь	314
10	з	408	22	н	982	34	ук гаммаподобный	225	46	ю	307
11	зело	133	23	о	1071	35	ф	280	47	юс малый	444
12	зело как s	10	24	о широкое	316	36	фита	171	48	ять	459

Табл. 4. Распределение символов в собранном датасете после аугментации

№	Символ	Кол-во	№	Символ	Кол-во	№	Символ	Кол-во	№	Символ	Кол-во
1	а	974	13	и палочка	335	25	омега	253	37	х	358
2	б	440	14	иже	1270	26	от	492	38	ц	286
3	в	872	15	ижица	281	27	п	547	39	ч	342
4	г	505	16	йотированная а	523	28	пси	102	40	ш	257
5	д	573	17	йотированная е	260	29	р	537	41	щ	267
6	е	520	18	к	428	30	с	1050	42	ъ	522
7	е как эпсилон	362	19	кси	207	31	т	714	43	ы с ъ	259
8	е широкое	251	20	л	411	32	у	281	44	ы с ь	381
9	ж	405	21	м	750	33	ук	329	45	ь	314
10	з	408	22	н	999	34	ук гаммаподобный	224	46	ю	307
11	зело	329	23	о	1071	35	ф	280	47	юс малый	444
12	зело как s	101	24	о широкое	299	36	фита	367	48	ять	458

Сведения об авторах

Егорова Аделя Марселевна, 1997 года рождения, аспирант 2-го года обучения в Национальном исследовательском ядерном университете «МИФИ» (НИЯУ МИФИ) по научной специальности 2.3.1 «Системный анализ, управление и обработка информации, статистика». Работает в НИЯУ МИФИ. Область научных интересов: компьютерное зрение, обработка изображений, классическое машинное обучение. E-mail: *AMKhasanova@mephi.ru*

Демидов Дмитрий Витальевич, 1980 года рождения. Высшее образование – специалитет: г. Москва, Московский инженерно-физический институт (государственный университет). 2003. Специальность «Прикладная математика». Квалификация «Инженер-математик». Ученая степень: кандидат технических наук. Доцент кафедры кибернетики (№22) института интеллектуальных кибернетических систем НИЯУ МИФИ. Область научных интересов: инженерия знания, обработка естественного языка, оптическое распознавание символов. E-mail: *DVDemidov@mephi.ru*

Егоров Алексей Дмитриевич, 1992 года рождения. Высшее образование – специалитет: ФГАОУ ВПО «Национальный исследовательский ядерный университет «МИФИ» г. Москва. 2014. Специальность «Прикладная математика и информатика». Квалификация «Математик, системный программист». Ассистент кафедры прикладной математики № 31 института лазерных и плазменных технологий НИЯУ МИФИ. Область научных интересов: компьютерное зрение, искусственный интеллект, хемоинформатика. E-mail: *ADEgorov@mephi.ru*

Поступила в редакцию 14 июня 2025 г. *Окончательный вариант – 09 декабря 2025 г.*

Classification of Church Slavonic Letter Images from the 10th–14th Centuries Using Neural Networks

A.M. Egorova [1], D.V. Demidov [1], A.D. Egorov [1]
[1]National Research Nuclear University MEPhI, Kashirskoye Shosse 31, Moscow, 115409, Russia

Abstract

This paper addresses a problem of recognizing Church Slavonic letters in texts from the 11th–15th centuries using machine learning methods. As part of the study, a database was created containing 20,180 images of 48 different letters, obtained from 647 spreads of liturgical books. To address class imbalance, data augmentation methods were applied.

Various neural network models were tested for letter classification, with EfficientNet-B4 achieving the best results, with F1-weighted scores of 0.976911. However, considering computational efficiency and processing time, an integral quality criterion was calculated, identifying EfficientNet-B1 as the optimal model for practical use. This model processes 43 characters per second, corresponding to 20 seconds per text spread.

The study results demonstrate the feasibility of efficient automated recognition of Church Slavonic texts, significantly accelerating their analysis and processing. Future work will focus on expanding the database and improving image preprocessing methods to enhance recognition accuracy.

Keywords: Artificial Intelligence, Image Classification, CNN.

Citation: Egorova AM, Demidov DV, Egorov AD. Classification of Church Slavonic Letter Images from the 11th–17th Centuries Using Neural Networks. Computer Optics 2026; 50(4): 1751. doi:10.18287/COJ1751.

Acknowledgements: This work was financially supported by the Ministry of Science and Higher Education Russian Federation within the "Priority-2030" NRNU MEPhI program.

Author's information

Adelya Marselevna Egorova, born in 1997, studied at the National Research Nuclear University MEPhI (NRNU MEPhI) as a second-year postgraduate student in the scientific specialty 2.3.1 «System Analysis, Control, and Information Processing, Statistics». She works at NRNU MEPhI. Her research interests include computer vision, image processing, and classical machine learning. E-mail: *AMKhasanova@mephi.ru*

Dmitry Vitalievich Demidov, born in 1980. Higher education – Specialist degree: Moscow, Moscow Engineering Physics Institute (State University), 2003. Specialty: «Applied Mathematics». Qualification: «Engineer-Mathematician». Academic degree: Candidate of Technical Sciences. Associate Professor at the Department of Cybernetics (No. 22) of the Institute of Intelligent Cybernetic Systems at NRNU MEPhI. Research interests: knowledge engineering, natural language processing, and optical character recognition. E-mail: *DVDemidov@mephi.ru*

Alexey Dmitrievich Egorov, born in 1992. Higher education – Specialist degree: National Research Nuclear University MEPhI, Moscow, 2014. Specialty: «Applied Mathematics and Informatics». Qualification: «Mathematician, Systems Programmer». Assistant at the Department of Applied Mathematics No. 31 of the Institute of Laser and Plasma Technologies at NRNU MEPhI. Research interests: computer vision, artificial intelligence, cheminformatics. E-mail: *ADEgorov@mephi.ru*

Received June 14, 2025. The final version – December 09, 2025.