

091(=163.41):004.352.243

<https://doi.org/10.18485/sj.2026.31.1.4>

ВЛАДИМИР Р. ПОЛОМАЦ*

ORCID: 0000-0001-5140-9330

Оригинални научни рад

Примљен: 5. 10. 2025.

Прихваћен: 15. 12. 2025.

МАРКО М. МИЛОШЕВИЋ

ORCID: 0009-0009-2402-7009

АНА МАРИЈА Б. ПАВЛОВИЋ

ORCID: 0009-0008-2558-4862

ТАМАРА Н. ЛУТОВАЦ КАЗНОВАЦ

ORCID: 0009-0003-0018-863X

Универзитет у Крагујевцу

Филолошко-уметнички факултет

КА ГЕНЕРИЧКОМ НТН МОДЕЛУ ЗА СРПСКЕ СРЕДЊОВЕКОВНЕ РУКОПИСЕ**

У раду је описан процес обуке и евалуације генеричког НТН (енгл. *Handwritten Text Recognition*) модела за аутоматско рашчитавање српских средњовековних рукописа писаних различитим типовима ћирилице (устав, полуустав и брзопис) помоћу софтверске платформе *Transkribus*. Најважнији резултат рада представља обука прве верзије робусног генеричког НТН модела *Мирослав 1.0* са изузетним перформансама – проценат погрешно рашчитаних карактера на скупу за проверу модела износи 3,53%. Евалуација перформанси овога НТН модела на

* v.polomac@filum.kg.ac.rs

** Рад је настао у оквиру међународног билатералног пројекта *Креирање AI модела за аутоматску обраду српских средњовековних рукописа*, који финансирају Министарство науке, технолошког развоја и иновација Републике Србије и Немачка служба за академску размену (DAAD). Претходна верзија рада саопштена је на међународној конференцији *Љужнословенски језици у дигиталном окружењу* (Филолошки факултет у Београду, 21–23. новембар 2024. године).

српским средњовековним рукописима који нису били део скупа података за обуку такође је показала изузетне резултате. Овај НТН модел веома успешно рашчитава српске средњовековне рукописе писане различитим типовима ћирилице, са процентом погрешно рашчитаних карактера који најчешће износи испод 10%.

Кључне речи: српски средњовековни рукописи, вештачка интелигенција, машинско учење, НТН (енгл. *Handwritten Text Recognition*), *Транскрибус*.

1. УВОД

Од пионирског рада А. Рабуса (2019) до данас област аутоматског рашчитавања словенских средњовековних рукописа помоћу вештачке интелигенције, машинског учења и напредних неуронских мрежа доживела је муњевит развој. Теоријски напредак огледа се у истраживању „паметних” функција НТН (енгл. *Handwritten Text Recognition*) модела (као што су разрешавање скраћеница или лигатура), затим у истраживању могућности корекције аутоматски добијених транскрипата помоћу великих језичких модела (LLM), као и интеграцији НТН технологије у теоријско-методолошки оквир квантитативне историјске корпусне лингвистике (уп. Рабус/Петров 2023, Поломац/Рабус 2024, Рабус 2024, Рабус/Журавељ 2025, Поломац 2025). На емпиријском плану напредак се највише огледа у развоју све ефикаснијих јавно доступних НТН модела помоћу софтверских платформи *Transkribus*¹ и *eScriptorium*² (Рабус 2022, Рабус/Томпсон 2023, Поломац 2023)³.

Први НТН модели за старо српско писано и штампано наслеђе развијени су пре неколико година за потребе аутоматског рашчитавања старих српских штампаних књига из Венеције (Поломац 2022а), а затим и књига из других српских штампарија XV–XVII века (Поломац 2022б). У основи оваког избора налазила се релативна уједначеност типографског слога старих српских штампаних књига XV–XVI века (у поређењу са честим варијацијама у словним облицима српских средњовековних рукописа) која гарантује успешну обуку НТН модела, чак и на мањим скуповима података. Након обуке генеричког НТН модела *Dionisio 2.0* за старе српске штампане књиге XV–XVI века са изузетним перформансама (процент погрешно препознатих карактера износио је само 2,4%) (Поломац 2022б) усмерили смо се на развој првих НТН модела за српске средњовековне рукописе. Избор је пао на рукописе Гаврила Стефановића Венцловића, једног од најзначајнијих српских књижевних стваралаца XVIII века, чија обимна ру-

¹ Детаљније в. на <https://app.transkribus.org/>.

² Детаљније о софтверској платформи *eScriptorium* в. у Кислинг и др. 2019.

³ Последња верзија генеричког НТН модела за црквенословенске рукописе *Generic Church Slavonic Handwriting 3* обучена је од стране А. Рабуса на огромном материјалу (око милион и по речи, углавном из руских црквенословенских рукописа, мање из старословенских рукописа). Детаљније о параметрима и перформансама овога модела в. на <https://www.transkribus.org/model/generic-church-slavonic-handwriting-3>.

кописна заоставштина обухвата више од двадесет рукописа са око десет хиљада листова (Поломац/Милошевић 2024: 292). Истраживање је најпре усмерено на обуку НТН модела за аутоматско рашчитавање Венцловићевих рукописа на српском језику за потребе историјског речника српског језика (Поломац и др. 2023), а затим и на обуку генеричког НТН модела за Венцловићеве српске и српскословенске рукописе писане полууставном и брзописном ћирилицом (Поломац/Милошевић 2024). Оба истраживања показала су се веома успешним, посебно друго, у оквиру кога је обучен генерички НТН модел са изузетним перформансама (процент погрешно рашчитаних карактера износио је 3%) (Поломац/Милошевић 2024: 296). Поређење овога НТН модела са генеричким НТН моделом за старе српске штампане књиге показало је да се и за рашчитавање српских средњовековних рукописа могу обучити ефикасни модели, само уз нешто већи скуп података за обуку.

У ослонцу на наведене резултате као циљ овога истраживања поставили смо обуку прве верзије робусног генеричког НТН модела за српске средњовековне рукописе писане различитим типовима средњовековне ћирилице (унав, полуунав и брзопис) помоћу софтверске платформе *Transkribus*.

2. ОБУКА ГЕНЕРИЧКОГ НТН МОДЕЛА ЗА СРПСКЕ СРЕДЊОВЕКОВНЕ РУКОПИСЕ

За обуку генеричког НТН модела искористили смо најпре скупове података припремљене за претходно обучене НТН моделе за рукописе Гаврила Стефановића Венцловића (модел *Венцл. 1.0* са процентом погрешно рашчитаних карактера од 3%, в. Поломац/Милошевић 2024: 296) и српске средњовековне повеље и писма (модел *Logothet 1.0* са процентом погрешно рашчитаних карактера од 5,6%, в. Поломац 2025). Скупови података за обуку наведених НТН модела приказани су у табелама 1 и 2.

Табела 1. Скуп података за обуку НТН модела за рукописе Гаврила Стефановића Венцловића

Рукопис	Тип писма, језик	Број речи
<i>Слова изабрана</i>	полуунав, српски	94 926
<i>Разглаголник</i>	брзопис, српскословенски	38 669
<i>Минеј за октобар</i>	полуунав, српскословенски	90 909
УКУПНО		224 504

Табела 2. Скуп података за обуку НТР модела за српске средњовековне повеље и писма

Рукопис	Тип писма	Број речи
<i>Повеље и писма XII–XVI века из Дубровника</i>	брзопис	56 388
<i>Дечанска хрисовуља (трећи примерак)</i>	полуустав	23 734
<i>Бањска хрисовуља</i>	устав	8 996
УКУПНО		89 118

Иако претходна два скупа података заједно садрже више од 300 000 речи, за обуку робусног генеричког НТР модела било је неопходно припремити још више података из средњовековних рукописа писаних уставом, полууставом и брзописом. Приликом одабира рукописа водили смо се потребом да се модел обучи на што разноврснијим подацима. Стога је следећи скуп података за обуку модела обухватио српскословенске и српске рукописе различитих жанрова писане уставом и полууставом из различитих хронолошких периода. Преглед рукописа из овога скупа приказан је у следећој табели.

Табела 3. Скуп података са српскословенским и српским рукописима

Рукопис	Тип писма, језик	Број речи
<i>Мирослављево јеванђеље</i> ⁴	устав, српскословенски	88 424
<i>Варлаам и Јоасаф</i> ⁵	полуустав, српскословенски	71 789
<i>Доментијаново Житије Св. Саве</i> ⁶	полуустав, српскословенски	60 700
<i>Студенички типик</i> ⁷	полуустав, српскословенски	21 506
<i>Стефанит и Ихнилат</i> ⁸	полуустав, српскословенски	18 669

⁴ Дигитални снимци преузети су са <https://kodeks.de/Serbia/MiroslavGospel.htm>.

⁵ Дигитални снимци рукописа *Варлаам и Јоасаф* (XIV век, манастир Хиландар, сигнатура Хил. 422) добијени су из Хиландарске истраживачке библиотеке у Колумбусу (Охајо, САД), на чему посебну захвалност изражавамо кустосу М. А. Џонсон, као и монасима манастира Хиландара.

⁶ Снимци Доментијановог Житија Св. Саве (препис, друга четвртина XIV века, Руска национална библиотека у Санкт Петербургу, сигнатура *Гљџф. 54*) преузети су са сајта пројекта *Српски рукописи у Руској националној библиотеци*: https://expositions.nlr.ru/ex_manus/Serbian_Manuscripts/serb/index.php.

⁷ Дигитални снимци рукописа преузети су из фототипског издања Анђелковић/Ракићевић 2019.

⁸ Снимци рукописа *Стефанит и Ихнилат* (препис, XVI век, Ћоровићева збирка Универзитетске библиотеке „Светозар Марковић” у Београду, сигнатура *Ћор. 31*) преузети су са сајта: <http://arhiva.unilib.rs/latinica/dokument/76/povedanije-o-drevnih-i-povjesteh-sa-stefanitom-i-ihnilatom>.

Душанов законик (Атонски препис) ⁹	полуустав, српски	7 709
Роман о Троју ¹⁰	полуустав, српски	10 951
УКУПНО		279 748

Процес рашчитавања свих рукописа из претходне табеле подразумевао је најпре ручно рашчитавање мањег броја страница, затим обучавање посебног НТR модела за сваки појединачни рукопис и аутоматско рашчитавање остатка текста, и на крају, ручно кориговање аутоматски рашчитаног текста.

На основу претходна три скупа података обучена је у софтверској платформи *Transkribus* прва верзија робусног генеричког НТR модела за српске средњовековне рукописе и старе штампане књиге под називом *Мирослав 1.0*¹¹. Параметри и перформансе овога модела приказани су у следећој табели.

Табела 4. *Параметри и перформансе генеричког НТR модела Мирослав 1.0*

Број речи	Број епоха	CER на скупу за обуку	CER на скупу за проверу
599 412	128	3,27%	3,53%

Наведена табела показује да је модел обучен на великом скупу података (скоро 600 000 речи) са изузетним перформансама – проценат погрешно рашчитаних карактера на скупу података за проверу¹² износи 3,53%. То практично значи да од сто карактера модел погрешно рашчитава само три или четири, укључујући и размак међу речима, знаке интерпункције, надредна слова и титле. Обука нове верзије генеричког модела на још разноврснијем скупу података вероватно не би значајније утицала на смањивање процента погрешно рашчитаних карактера на скупу за проверу модела, али би свакако утицала на његове перформансе у рашчитавању рукописа који нису били део скупа података за обуку.

⁹ Дигитални снимци преузети су са сајта Руске државне библиотеке у Москви: <https://libfond.ru/lib-rgb/87/f-87-28/#image-241>.

¹⁰ Преузето са сајта Народне библиотеке Св. Ђирило и Методије у Софији: <https://digilib.nationallibrary.bg/slr/public/image>.

¹¹ Модел је назван у част *Мирослављевог јеванђеља*, најрепрезентативнијег српског средњовековног рукописа.

¹² Скуп података за обуку аутоматски је подељен тако да је 2% података употребљено за проверу (испитивање перформанси) модела.

3. КВАНТИТАТИВНА ЕВАЛУАЦИЈА ГЕНЕРИЧКОГ НТР МОДЕЛА *МИРОСЛАВ 1.0*

Квантитативне перформансе НТР модела *Мирослав 1.0* тестиране су кроз три експеримента. Због жеље да тестирамо перформансе НТР модела на што разноврснијем корпусу, у свим експериментима ограничили смо се само на по једну страницу из сваког рукописа.

Првим експериментом обухваћено је неколико српских и српскословенских рукописа Гаврила Стефановића Венцловића који нису били део скупа за обуку. У табели 5 приказани су резултати поређења квантитативних перформанси модела *Мирослав 1.0* и *Венцл. 1.0*.

Табела 5. Евалуација НТР модела на рукописима
Гаврила Стефановића Венцловића

Рукопис	Тип писма, језик	<i>Венцл. 1.0</i>	<i>Мирослав 1.0</i>
<i>Минеј за новембар</i>	полуустав, српскословенски	4,71%	5,56%
<i>Великописник</i>	полуустав, српски	4,99%	5,98%
<i>Пресађеница</i>	брзопис, српскословенски	3,55%	10,04%
<i>Поученије изабрано</i>	полуустав, српски	2,60%	3,14%

Наведена табела указује на сличне перформансе двају модела у рашчитавању Венцловићевих српскословенских рукописа писаних полууставом. Нешто слабије перформансе модела *Мирослав 1.0* представљају резултат ширења скупа података за обуку другим српскословенским рукописима писаним полууставом. Поређење двају модела у рашчитавању Венцловићевих српскословенских рукописа писаних брзописом открива значајне разлике у перформансама. Модел обучен само на Венцловићевим рукописима показује значајно боље резултате од генеричког модела *Мирослав 1.0*. Претпостављамо да је овакав резултат последица значајно мањег обима Венцловићевих српскословенских рукописа писаних брзописом у скупу података за обуку генеричког модела *Мирослав 1.0*. Из овога експеримента јасно проистиче закључак да је за наставак рада на рашчитавању Венцловићевих рукописа боље користити модел обучен само на његовим рукописима.

Другим експериментом обухваћено је неколико српских средњовековних повеља и писама који нису били део скупа за обуку модела *Мирослав 1.0*. И овде су квантитативни показатељи евалуације упоређени са резултатима НТР модела *Logothet 1.0* (в. табелу 6).

Табела 6. Евалуација НТН модела на српским средњовековним повељама и писмима

Рукопис	Тип писма	<i>Logothet 1.0</i>	<i>Мирослав 1.0</i>
<i>Повеља монахиње Јевђеније Великој Лаври</i> ¹³	брзопис	15,09%	10,88 %
<i>Дубровачко писмо војводи Сандаљу</i> ¹⁴	брзопис	4,43%	3,24%
<i>Повеља деспота Стефана Великој Лаври</i> ¹⁵	полуустав	10,30%	4,45%

Наведена табела показује најпре како генерички модел *Мирослав 1.0* постиже значајно боље резултате у рашчитавању повеља и писама од модела *Logothet 1.0*. Овакав резултат вероватно проистиче из разлике у величини и саставу скупа података за обуку двају модела. Аутоматско рашчитавање повеља и писама (посебно манастирских даровница) значајно је унапређено укључивањем других српскословенских рукописа у скуп података за обуку модела. Из овога експеримента проистиче закључак како се наставак рада на аутоматском рашчитавању српских средњовековних повеља и писама може најефикасније наставити помоћу генеричког модела *Мирослав 1.0*.

Последњим експериментом обухваћено је неколико српскословенских рукописа различитих жанрова насталих од XIII до XVIII века. Резултати евалуације приказани су у следећој табели.

Табела 7. Евалуација НТН модела *Мирослав 1.0* на српскословенским рукописима

Рукопис	Тип писма	CER
<i>Вуканово јеванђеље</i> (XIII век) ¹⁶	устав	12,93%
<i>Матичин апостол</i> (XIII век) ¹⁷	устав	8,41%
<i>Минеј за јул</i> (XIV век) ¹⁸	полуустав	9,44%

¹³ Више о овој повељи у Младеновић 2007: 223–228.

¹⁴ Више о писму в. у Стојановић 1929: 286–287.

¹⁵ Више о овој повељи у Младеновић 2007: 253–257.

¹⁶ Дигитални снимци рукописа преузети су са сајта Руске националне библиотеке: <https://nlr.ru/manuscripts/RA1527/elektronnyiy-katalog?ab=B6787682-05A5-4EFE-A251-2FBVBF954930>.

¹⁷ Снимци рукописа преузети су са сајта Дигиталне библиотеке Матице српске: <https://digital.bms.rs/ebiblioteka/webroot/index.php?url=publications/view/743>.

¹⁸ Дигитални снимци рукописа преузети су са сајта Универзитетске библиотеке „Светозар Марковић” у Београду: <http://www.ubsm.bg.ac.rs/cirilica/dokument/61/minej-za-jul>.

<i>Паренезис Јефрема Сирина</i> (XVI век) ¹⁹	полуустав	6,88%
<i>Псалтир</i> Гаврила Тројичанина (XVII век) ²⁰	полуустав	5,45%
<i>Стихологија</i> Кипријана Рачанина (XVIII век) ²¹	полуустав	4,58%

На основу наведене табеле може се закључити како су перформансе НТР модела *Мирослав 1.0* у рашчитавању српскословенских рукописа одличне, будући да проценат погрешно рашчитаних карактера најчешће износи мање од 10%. Лошији резултат у рашчитавању *Вукановог јеванђеља* не представља изузетак пошто се највећи број грешака односи на погрешно рашчитавање двотачке која је веома фреквентна у овом рукопису, а није довољно присутна у скупу података на коме је модел обучен.

4. ЗАКЉУЧНЕ НАПОМЕНЕ

Проведено истраживање показало је да се помоћу софтверске платформе *Transkribus*, засноване на принципима вештачке интелигенције, машинског учења и напредних неуронских мрежа, може креирати изузетно ефикасан генерички (општи) модел за аутоматско рашчитавање српских средњовековних рукописа писаних различитим типовима ћирилице (устав, полуустав и брзопис). Најважнији практични резултат рада представља обука прве верзије генеричког модела за српске средњовековне рукописе названог *Мирослав 1.0*, у част Мирослављевог јеванђеља, најрепрезентативнијег српског средњовековног споменика. Овај модел је обучен на великом и разноврсном скупу података који чини око 600 000 речи из средњовековних рукописа XII–XVIII века различитих жанрова писаних различитим типовима ћирилице. Квантитативна евалуација перформанси модела показала је да проценат погрешно рашчитаних карактера на рукописима који нису били део скупа за обуку најчешће варира 5–10%, што се може оценити одличним. Помоћу овога модела превођење текста српских средњовековних рукописа у машински читљив облик вишеструко ће се убрзати чиме ће се отворити нове перспективе за квантитативно и корпусно заснована истраживања старог српског писаног наслеђа. Од посебног значаја јесте чињеница да се модел даље може усавршавати ширењем скупа података за обуку, чиме се његове перформансе могу додатно унапредити. У

¹⁹ Дигитални снимци рукописа преузети су са сајта Копитареве збирке Универзитетске библиотеке у Љубљани: <https://www.dlib.si/details/URN:NBN:SI:IMG-ALOB5JG5>.

²⁰ Преузето са сајта Дигиталне библиотеке Матице српске: <https://digital.bms.rs/ebiblioteka/publications/view/773>.

²¹ Дигитални снимци рукописа преузети су са сајта Дигиталне библиотеке Матице српске: <https://digital.bms.rs/ebiblioteka/webroot/index.php?url=publications/view/2323>.

наредном периоду модел ће остати приватан, уз могућност бесплатног уступања на захтев корисницима софтверске платформе *Transkribus*. Поред даљег усавршавања модела ширењем скупа података за обуку, у блиској будућности планирамо и обуку модела помоћу софтверске платформе *eScriptorium* како би заинтересовани корисници могли имати бесплатну алтернативу софтверској платформи *Transkribus*.

ЛИТЕРАТУРА

- Анђелковић/Ракићевић 2018:** М. Анђелковић, Т. Ракићевић, *Студенички типик. The Studenica Turikon. Студенички типикон*, Студеница: Манастир Студеница.
- Кислинг и др. 2019:** В. Kiesling, R. Tissot, P. Stokes, D. Stökl Ben Ezra, *eScriptorium: An Open Source Platform for Historical Document Analysis*, in: *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)* (pp. 19–24). <https://ieeexplore.ieee.org/document/8893029> <13. 6. 2025>
- Младеновић 2007:** А. Младеновић, *Повеље и писма деспота Стефана*, Београд: Чигоја штампа.
- Поломац 2022а:** V. Polomac, Serbian Early Printed Books from Venice: Creating Models for Automatic Text Recognition using *Transkribus*, *Scripta & e-Scripta*, 22, 11–29.
- Поломац 2022б:** V. Polomac, Serbian Early Printed Books: Towards Generic Model for Automatic Text Recognition using *Transkribus*, in: T. Erjavec, D. Fišer (eds.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika/Proceedings of the Conference on Language Technologies and Digital Humanities*, 154–161.
- Поломац 2023:** V. Polomac, *Macarius*: A HTR model for Romanian Slavonic Early Printed Books, *Slavistica Vilnensis*, 68/2, 10–23.
- Поломац и др. 2023:** V. Polomac, М. Kurešević, I. Bjelaković, А. Colić Jovanović, S. Petrović, Digitizing Cyrillic Manuscripts for the Historical Dictionary of the Serbian Language Using Handwritten Text Recognition Technology, *Slověne: International Journal of Slavic Studies*, 12/1, 295–316.
- Поломац/Милошевић 2024:** В. Поломац, М. Милошевић, Ка (дигиталном) издању рукописа Гаврила Стефановића Венцловића помоћу вештачке интелигенције, *Књижевност и језик*, 71/2, 291–302.

- Поломац/Рабус 2024:** V. Polomac, A. Rabus, Serbian Early Printed Books from Venice: A Quantitative Approach to Orthographic Variations, *Studi Slavistici*, XXI/2, 37–60.
- Поломац 2025:** В. Поломац, Теоријско-методолошки оквир за изградњу електронског корпуса српских средњовековних повеља и писама, у: Р. Драгићевић, Б. Сувајдић (ур.), *Српска славистика 2025*, 177–198.
- Рабус 2019:** A. Rabus, Recognizing handwritten text in Slavic manuscripts: A neural network approach using *Transkribus*, *Scripta&e-Scripta*, 19, 9–32.
- Рабус 2022:** A. Rabus, Handwritten Text Recognition for Croatian Glagolitic, *Slovo*, 72, 181–192.
- Рабус/Петров 2023:** A. Rabus, I. Petrov, Linguistic Analysis of Church Slavonic Documents: A Mixed-Methods Approach, *Scando-Slavica*, 69/1, 25–38.
- Рабус/Томпсон 2023:** A. Rabus, W. Thompson, Performance of Generic HTR Models on Historical Cyrillic and Glagolitic: Comparison of Engines, *Scripta & e-Scripta*, 23, 11–34.
- Рабус 2024:** A. Rabus, Tolerating Imperfection: Uncorrected Transkribus Transcriptions in Church Slavonic Studies, in: L. Taseva, A. Rabus, I. Petrov (eds.), *Učitelnoto evangelie na Konstantin Preslavski i južno-slavjanskite prevodi na homiletični tekstove (IX–XIII v.)*, 453–467.
- Рабус/Журавель 2025:** A. Rabus, A. Jouravel, Handwritten Text Recognition and the Future of Paleoslavonic Studies, in: D. Bunčić (ed.), *Deutsche Beiträge zum 17. Internationalen Slavistikongress Paris 2025, Resümees*, 217–218.
- Стојановић 1929:** Љ. Стојановић, *Старе српске повеље и писма, књ. I*, Београд: Српска краљевска академија.
- Трифуновић 2011:** Ђ. Трифуновић (ур.), *Повеља краља Милутина манастиру Бањска. Светостефанска хрисовуља*, Београд: Службени гласник – Центар за очување наслеђа Косова и Метохије *Mnemosyne*.

TOWARDS GENERIC HTR MODEL FOR SERBIAN MEDIEVAL MANUSCRIPTS

Summary

This paper presents the process of training and evaluating a general-purpose HTR (Handwritten Text Recognition) model designed for the automatic transcription of Serbian medieval manuscripts written in various forms of Cyrillic script, utilizing the *Transkribus* software platform. The primary practical outcome of this research is the development of the first version of a generic HTR model for Serbian medieval manuscripts, named *Miroslav 1.0*, in honor of the *Miroslav Gospel*, the most representative manuscript of the Serbian medieval tradition. The model was trained on a large and heterogeneous dataset comprising approximately 600,000 words extracted from 12th to 18th century manuscripts of various genres, all written in distinct styles of Cyrillic script (uncial, semi-uncial and cursive). The quantitative evaluation of the model's performance indicates a character error rate (CER) ranging between 5% and 10% on out-of-sample manuscripts, which is considered highly satisfactory for historical manuscript transcription tasks. The implementation of this model significantly accelerates the transcription process of Serbian medieval manuscripts into machine-readable formats, thereby opening new avenues for corpus-based and quantitative research into the Serbian written heritage. Particularly noteworthy is the model's extensibility: its accuracy and robustness can be further enhanced by expanding the training dataset with additional material. In addition to future improvements based on dataset augmentation, we also plan to train the model using the *eScriptorium* platform. This will provide researchers with an open-access alternative to *Transkribus*, thereby promoting broader accessibility and sustainability of digital tools in Slavic medieval manuscript studies.

Keywords: Serbian Medieval Manuscripts, Artificial intelligence, machine learning, HTR (Handwritten Text Recognition), *Transkribus* software platform.

Vladimir R. Polomac
Marko M. Milošević
Ana Marija B. Pavlović
Tamara N. Lutovac Kaznovac