

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Оцена научне заснованости теме докторске дисертације кандидата Лазара Смиљковића, мастер инжењера електротехнике и рачунарства

Одлуком Наставно-научног већа Електротехничког факултета бр. 1120/12 од 27.06.2025. године, именовани смо за чланове Комисије за научне заснованости теме докторске дисертације кандидата Лазара Смиљковића, мастер инжењера електротехнике и рачунарства, за израду докторске дисертације и научне заснованости теме **„Развој и евалуација метода за анализу и унапређење квалитета података просторне транскриптомике“** (енг. *“Development and evaluation of methods for analyzing and improving the quality of spatial transcriptomics data”*).

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

ИЗВЕШТАЈ

1. Подаци о кандидату

1.1. Биографски подаци

Лазар Смиљковић рођен је 11.04.1998. у Крушевцу, Република Србија. Основну школу „Миодраг Чајетинац Чајка“ и гимназију „Вук Караџић“ завршио је у Трстенику, оба пута као носилац Вукове дипломе и ученик генерације. Током школовања освојио је дванаест награда на државним такмичењима из математике, физике и информатике.

Након завршене средње школе, 2017. године уписао је Електротехнички факултет Универзитета у Београду, студијски програм Софтверско инжењерство. Дипломирао је 2021. године са просеком 9,93. Дипломски рад, под називом „Реализација веб система за квиз Стигни ме ако знаш“ и под руководством др Дражена Драшковића, одбранио је са оценом 10.

Дипломске академске – мастер студије уписао је 2021. године на Електротехничком факултету у Београду, модул Софтверско инжењерство. Мастер студије је завршио 2022. године са просечном оценом 10,00. Завршни рад, под називом „Софтверски модул за анализу изворних података о комуникационим активностима у рачунарским мрежама“ и под руководством др Славка Гајина, одбранио је са оценом 10.

Докторске академске студије уписао је 2022. године на Електротехничком факултету у Београду, модул Софтверско инжењерство, под менторством др Марка Мишића. Положио

све испите са просечном оценом 10,00. Коаутор је на раду у часопису са SCI листе и објавио је 7 научних радова на домаћим и страним конференцијама као аутор или коаутор.

Током студирања је три године био ангажован као студент демонстратор на више предмета са Катедре за рачунарску технику и информатику. Од 2022. године запослен је као *full stack* програмер у фирми Sonесо, а од јануара 2023. је запослен као истраживач приправник на Електротехничком факултету Универзитета у Београду. Током лета 2024. године боравио је у Кини на „2024 BGI International Summer Workshop & ICG 19 Youth Symposium“ где је и презентовао резултате својих истраживања. Био је учесник на међународном пројекту „Razvoj novih algoritama за OMICS podatke (ETF-OMICS)“ 2023 2024. године током ког је сарађивао са колегама из страних и домаћих фирми у развоју алгоритама за обраду података из области просторне транскриптомике.

1.2. Сачењено научноистраживачко искуство

Лазар Смиљковић је научно-истраживачко искуство започео 2022. године уписом на докторске академске студије на Електротехничком факултету Универзитета у Београду, за област Рачунарска техника и информатика, и на тим студијама положио је следеће испите:

- | | |
|---|------------------|
| - Принципи програмских језика (9 ЕСПБ) | оцена 10 (десет) |
| - Формални језици и аутомати (9 ЕСПБ) | оцена 10 (десет) |
| - ТЦП/ИП архитектура (9 ЕСПБ) | оцена 10 (десет) |
| - Напредни алгоритми машинског учења (9 ЕСПБ) | оцена 10 (десет) |
| - Развој микропроцесорског софтвера (9 ЕСПБ) | оцена 10 (десет) |
| - Архитектуре за подршку софтверским системима (9 ЕСПБ) | оцена 10 (десет) |
| - Увод у научни рад (6 ЕСПБ) | оцена 10 (десет) |

Поред положених испита, кандидат је испунио и све остале обавезе предвиђене наставним планом и програмом: Студијски истраживачки рад I (24 ЕСПБ), Студијски истраживачки рад II (24 ЕСПБ) и Научно стручни рад (3ЕСПБ).

Од 2022. године кандидат је учествовао на једном научном пројекату на Електротехничком факултету Универзитета у Београду:

- 1.1. Л. Смиљковић (учесник), М. Мишић (руководилац), Razvoj novih algoritama за OMICS podatke (ETF-OMICS), на Електротехничком факултету Универзитета у Београду, финансираног од стране COMPLETE GENOMICS Inc., пројекат број 1847, 2022-2024.

Лазар Смиљковић је аутор 8 научних публикација. Коаутор је једног рада у научном часопису са *impact* фактором. У зборницима радова међународних скупова је објавио три научна рада, док је на скуповима националног значаја објавио 4 рада.

Главне области интересовања су му биоинформатика и машинско учење. Нарочиту пажњу посвећује истраживању и развоју алгоритама за обраду биолошких података, а нарочито података просторне транскриптомике.

Публикације на научним скуповима

Међународни часописи са *impact* фактором

- 1.1. S. Branković, **L. Smiljković**, P. Obradović, M. Radonjić, M. Mišić, Fast Parallel CPU-GPU Approximate Spectral Clustering for Transcriptomics Data, INTERNATIONAL JOURNAL OF PARALLEL PROGRAMMING, Vol. 53, No. 5, pp. 1 - 25, Jan, 2025. [ISSN: 1573-7640, M23, IF2024=0.9]

Међународне конференције

- 1.1. **L. Smiljković**, M. Radonjić, M. Mišić, M. Tomašević, Comparing Python Code Parallelization Techniques for Spatial Transcriptomics Data, XXXI Telecommunications forum Telfor 2023, Društvo za telekomunikacije, Belgrade, Nov, 2023 [M33]
- 1.2. **L. Smiljković**, P. Obradović, M. Mišić, V. Kovačević, Incorporating Practical Single Cell and Spatial Transcriptomics Analysis in a Bioinformatics Course, 2024 11th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), IEEE, Nis, Serbia, Jun, 2024 [M33]
- 1.3. M. Mišić, **L. Smiljković**, P. Obradović, Spectral Clustering for Transcriptomics Data: an Approximate Column Sampling Approach on the GPU, 5th Belgrade Bioinformatics Conference – BelBi 2024, Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Belgrade, Serbia, Jun, 2024 [M34]

Домаће конференције

- 1.1. **Л. Смиљковић**, М. Мишић, П. Вулетић, С. Гајин, Софтверски алат за анализу изворних података о комуникационим активностима у рачунарским мрежама, Зборник радова, LXVII конференција ЕТРАН, Друштво за ЕТРАН, Источно Сарајево, Босна и Херцеговина, Јун, 2023 [M63]
- 1.2. **Л. Смиљковић**, М. Мишић, Д. Драшковић, Упоредивање модела машинског учења на примеру биоинформатичких података, Зборник радова 30. ИКТ конференције "YU INFO 2024", pp. 187 - 191, Информационо друштво Србије, Копаоник, Србија, Мар, 2024 [M63]
- 1.3. **Л. Смиљковић**, М. Мишић, В. Ковачевић, Евалуација статистичких метода импутације гена у подацима просторне транскриптомике, Зборник радова 31. ИКТ конференције "YU INFO 2025", pp. 187 - 191, Информационо друштво Србије, Копаоник, Србија, Мар, 2025 [M63]
- 1.4. **Л. Смиљковић**, М. Мишић, Ј. Протић, Мастер студијски програми из биоинформатике у Србији и свету, Зборник радова, LXIX конференција ЕТРАН, Друштво за ЕТРАН, Чачак, Србија, Јун, 2025 [M63]

Дана 11.07.2025. године одржана је јавна усмена одбрана теме докторске дисертације Лазара Смиљковића, пред комисијом коју су чинили:

- др Јелица Протић, редовни професор Електротехничког факултета Универзитета у Београду,
- др Владимир Ковачевић, научни сарадник, Истраживачко-развојни институт за вештачку интелигенцију Србије,
- др Милица Јанковић, ванредни професор Електротехничког факултета Универзитета у Београду.

Комисија је кандидата на јавној усменој одбрани оценила оценом *задовољно*.

1.3. Оцена подобности кандидата за рад на предложеној теми

Комисија је размотрила све претходно наведене резултате досадашњег рада кандидата Лазара Смиљковића. Кандидат је кроз положене испите на докторским студијама, остварено учешће у стручним пројектима, успешну одбрану теме докторске дисертације и до сада објављене научне радове у часописима и на научно-стручним конференцијама, показао да је стекао неопходно искуство за самосталан научно-истраживачки рад. Комисија сматра да је кандидат способан да током израде докторске дисертације оствари резултате који ће бити верификовани од стране међународне научне заједнице. Комисија констатује да кандидат Лазар Смиљковић испуњава све потребне услове да приступи изради докторске дисертације под наведеном темом.

2. Предмет, циљ и значај истраживања

Генско секвенцирање представља поступак читања и дешифровања редоследа нуклеотида, основних јединица ДНК, у геному одређеног организма. Ова технологија чини темељ савремене биологије и медицине, јер омогућава детаљно разумевање структуре, функције и регулације гена кроз проучавање ДНК и РНК молекула. Подаци добијени генским секвенцирањем пружају увид у то који су гени активни, у којој мери је њихова експресија присутна, као и како се она мења под различитим физиолошким или патолошким условима, а област која проучава РНК секвенце се назива транскриптомика. Један од значајних помака у овој области остварен је развојем технологије секвенцирања појединачних ћелија (енг. single-cell RNA sequencing – scRNA-seq). Ова метода омогућава прецизно мерење експресије гена у свакој ћелији појединачно, чиме се открива ћелијска разноликост, идентификују нови типови ћелија, прати њихова динамика током развоја организма, као и током различитих болести. Једно од главних ограничења scRNA-seq технологије јесте то што се приликом изоловања и обраде ћелија губи просторни контекст, односно не остаје сачувана информација о томе где се у ткиву одређена ћелија првобитно налазила.

Управо тај недостатак превазилази област истраживања, заснована на новијим технологијама, просторна транскриптомика. Она омогућава не само анализу тога који су гени активни, већ и одређивање њихове експресије у конкретним деловима ткива. Уобичајена процедура подразумева сечење ткива на танке пресеке који се постављају на специјалне носаче са унапред дефинисаним просторним бар-кодovima. Свака позиција на носачу има своју „адресу“, што омогућава да се после секвенцирања добијени транскрипти прецизно повежу са њиховим положајем у ткиву. Оваква информација је од изузетног значаја за разумевање биолошких процеса који укључују сложене међућелијске интеракције. Просторна транскриптомика омогућава, између осталог, проучавање утицаја ћелија једне на

другу у развоју органа, ширењу туморских ћелија, имунолошким одговорима и поремећајима нервног система.

Иако просторна транскриптомика представља значајан технолошки напредак, она и даље има одређена технолошка ограничења која утичу на квалитет и интерпретабилност добијених података. Један од најзначајнијих проблема је тзв. dropout ефекат, појава у којој многи гени, иако присутни и активни у ћелији, нису детектовани у подацима. Ово доводи до вештачког губитка информација, односно до тога да се неки гени погрешно тумаче као неекспресовани. Ови технички недостаци значајно утичу на даљу анализу и интерпретацију података. Разлози за овај ефекат су техничке природе и укључују малу дубину секвенцирања, ограничену осетљивост метода и деградацију РНК током обраде узорака. Осим тога, просторна резолуција неких метода није довољно висока да би се постигла ћелијска резолуција, већ се добијају просеци експресије више ћелија у једној тачки.

Један од најперспективнијих начина да се превазиђу наведена ограничења јесте интеграција просторне транскриптомике са подацима добијеним из scRNA-seq технологије. Док scRNA-seq нуди високу резолуцију и осетљивост на нивоу појединачне ћелије, просторна транскриптомика додаје контекст, односно географски положај ћелија у ткиву. Комбиновањем ових података могуће је импутовати недостајуће експресије гена у просторним подацима чиме се њихов квалитет и тачност побољшавају.

Предмет истраживања ове докторске дисертације су анализа и поређење постојећих метода и развој нових метода за интеграцију података просторне транскриптомике и scRNA секвенцирања, са посебним фокусом на технике које служе за импутацију недостајућих вредности у просторним подацима. Истраживање обухвата систематичан преглед постојећих решења као што су Tangram, Seurat и други. Посебна пажња посвећена је идентификовању недостатака ових алата, укључујући осетљивост на квалитет улазних података, пренаглашену зависност од анотације ћелијских типова и недовољну робусност на техничке варијације.

Циљ истраживања је да се на основу претходне анализе предложи и реализују одговарајућа метода и радни оквир за импутацију недостајућих вредности у просторној транскриптомици, који користи упарене scRNA-seq податке са истих или сродних узорака. Идеја је да се избегну и реше недостаци других решења која често не пружају довољно добру тачност, нису робусна на различите типове података и техничке варијације, или захтевају комплексну и скупу припрему података. У циљу свеобухватне и репрезентативне евалуације предложеног приступа, у раду ће се користити више скупова података различитих карактеристика. Одабрани скупови обухватају различите типове ткива, различите технологије просторне транскриптомике и различите величине, од мањих скупова са стотинама тачака до великих са стотинама хиљада ћелија. Таква разноврсност омогућава тестирање алата у реалистичним условима, као и процену његове робусности, флексибилности и применљивости у различитим биолошким контекстима. Резултати ће бити евалуирани кроз метрике тачности, очување биолошких сигнала и стабилност у различитим условима. Додатно, фокус ће бити и на побољшању перформанси уз разматрање техника паралелизације како би се искористили пуни потенцијали рачунарског хардвера на коме се врши обрада.

Значај истраживања је у унапређењу квалитета података у области просторне транскриптомике, кроз развој метода за исправку недостајућих вредности које настају као последица техничких ограничења. Побољшањем квалитета података добија се прецизнија и поузданија слика просторне организације експресије гена, што директно утиче на квалитет и вредност биолошких анализа које се на тим подацима заснивају. Истраживање такође доприноси бољем разумевању тренутних ограничења постојећих алата, као и начина на који

се та ограничења могу превазићи комбинацијом различитих извора података и применом напредних аналитичких техника. На тај начин, предлог решења који настаје као резултат рада представља не само технички, већ и методолошки допринос области. Осим академске вредности, резултати овог истраживања имају и ширу примену у научним и индустријским окружењима, посебно у областима као што су биомедицина, фармакогеномика, онкологија, неуронауке и развој персонализованих терапија. Унапређени алати за анализу просторних података могу допринети бољем разумевању сложених биолошких процеса и отворити пут ка прецизнијим дијагностичким и терапијским приступима.

Најрелевантније референце у вези са истраживањем

- [1] Moses, L., & Pachter, L. (2022). Museum of spatial transcriptomics. *Nature methods*, 19(5), 534-546.
- [2] Saliba, A. E., Westermann, A. J., Gorski, S. A., & Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. *Nucleic acids research*, 42(14), 8845-8860.
- [3] Zeng, Z., Li, Y., Li, Y., & Luo, Y. (2022). Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome biology*, 23(1), 83.
- [4] Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, 33(5), 495-502.
- [5] Achim, K., Pettit, J. B., Saraiva, L. R., Gavriouchkina, D., Larsson, T., Arendt, D., & Marioni, J. C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nature biotechnology*, 33(5), 503-509.
- [6] Li, Y., Stanojevic, S., & Garmire, L. X. (2022). Emerging artificial intelligence applications in spatial transcriptomics analysis. *Computational and Structural Biotechnology Journal*, 20, 2895-2908.
- [7] Wan, X., Xiao, J., Tam, S. S. T., Cai, M., Sugimura, R., Wang, Y., ... & Yang, C. (2023). Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nature Communications*, 14(1), 7848.
- [8] Villani, C. (2008). Optimal transport: old and new (Vol. 338, pp. 129-131). Berlin: springer.
- [9] Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40, 99-121.
- [10] Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D. J., Hicks, S. C., Robinson, M. D., ... & Schönhuth, A. (2020). Eleven grand challenges in single-cell data science. *Genome biology*, 21, 1-35.
- [11] Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143(143), 8.
- [12] Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., ... & Regev, A. (2021). Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nature methods*, 18(11), 1352-1362.
- [13] Little, R. J., & Rubin, D. B. (2019). Statistical analysis with missing data. John Wiley & Sons.
- [14] Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artificial intelligence in medicine*, 50(2), 105-115.
- [15] Candes, E., & Recht, B. (2012). Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6), 111-119.

3. Полазне хипотезе

Основне хипотезе од којих се полази у овом истраживању су:

- Просторна информација игра кључну улогу у анализи транскриптомских података, јер омогућава боље разумевање ћелијског контекста, интеракција и функције ткива
- Комбиновањем података просторне транскриптомике са scRNA-seq подацима могуће је значајно умањити утицај dropout ефекта и побољшати тачност експресионих профила.
- Анализом и упоређивањем више постојећих алата за импутацију могу се идентификовати кључне слабости тренутних приступа и области у којима постоји простор за унапређење.
- Импутација недостајућих вредности у матрицама података просторне транскриптомике може се формулисати као оптимизациони проблем, што омогућава примену различитих алгоритамских техника ради побољшања резултата.
- Матрице генске експресије су структурално погодне за паралелну обраду, што омогућава убрзање обраде применом савремених паралелних техника обраде података.
- Аутоматизован и модуларно дизајниран радни оквир за импутацију може омогућити лакше тестирање, поређење и примену у различитим биолошким и технолошким контекстима, чиме се повећава његова научна и практична вредност.

4. Научне методе истраживања

Уводна фаза ће обухватити систематичан преглед релевантне научне литературе, са фокусом на најновија достигнућа у области просторне транскриптомике и њене интеграције са scRNA-seq подацима. Посебна пажња биће посвећена методама импутације, начинима решавања dropout ефекта, као и процени перформанси постојећих алата. За потребе анализе биће коришћени скупови података различитих величина, из различитих биолошких ткива и технологија просторног секвенцирања (нпр. 10x Genomics Visium, Slide-seq, MERFISH). Подаци обухватају и просторне и упарене scRNA-seq податке, што омогућава свеобухватно тестирање и поређење са другим приступима. Ново решење биће упоређено са већ постојећима користећи више метрика као што су тачност, ретенција маркер гена и визуелна процена – поређење слика ткива пре и после импутације.

С обзиром на то да *dropout* ефекат у просторним подацима доводи до појаве великог броја нултих вредности, имплементирана решења не смеју нарушавати основну дистрибуцију података. Након завршене импутације потребно је да структура *single cell* матрица (енг. sparsity) остане очувана. У другом кораку истраживања биће тестиране једноставније статистичке методе на основу дистрибуције генских експресија.

У наставку истраживања биће дискутован и формиран одговарајући оптимизациони модел који одговара проблему који се решава. Циљ је поклапање експресионих дистрибуција сваког гена у просторном ткиву са дистрибуцијама у scRNA подацима. Биће тестиране различите мере растојања између дистрибуција, као што су Kullback-Leibler дивергенција и Earth Mover's Distance. Додатно, на проблем ће бити примењени и алгоритми машинског учења као што је КНН, у циљу импутације нових вредности. Алгоритми ће бити тренирани на scRNA подацима и тестирани на просторним матрицама уз прилагођавање архитектуре модела природе проблема. Сви резултати ће бити валидарани на неколико скупова података

различитог порекла, а перформансе система поређене са резултатима добијеним из постојећих алата.

5. Очекивани научни допринос

Ова дисертација има за циљ да пружи конкретан научни и применљив допринос области анализе просторне транскриптомике, са фокусом на проблем импутације података коришћењем упарених scRNA-seq података. Очекује се да истраживање доведе до следећих резултата и научних доприноса:

- Систематизована и критичка анализа постојећих алата и метода за импутацију генских експресија у податке просторне транскриптомике.
- Моделовање проблема импутације као оптимизационог задатка, са циљем минимизовања разлика између просторне и референтне (scRNA) експресионе дистрибуције, користећи релевантне метрике растојања између дистрибуција.
- Имплементација решења у виду софтверског радног оквира са фокусом на ефикасности и могућности паралелног извршавања, посебно у контексту обраде великих експресионих матрица.
- Евалуација имплементираних система у смислу тачности импутације, очувања биолошки значајних сигнала и укупне рачунарске ефикасности (време извршавања, меморијски захтеви), уз поређење са најрепрезентативнијим постојећим решењима.
- Евалуација новог решења добијена на основу различитих скупова података који се разликују по врсти ткива, технологији и величини.

6. План истраживања и структура рада

Планирано је да се истраживање спроведе у следећим фазама:

- У првој фази биће спроведена свеобухватна анализа постојећих алата и метода за импутацију у просторној транскриптомици. Поред функционалности, посебна пажња биће посвећена метрикама евалуације, коришћеним моделима, типовима података и ограничењима постојећих приступа.
- У другој фази, на одабраним скуповима података биће примењени једноставни статистички алгоритми са циљем да се испита базична могућност импутације уз очување структуре података, нарочито у погледу броја нултих вредности.
- Трећа фаза предвиђа конструисање формалног модела проблема импутације као оптимизационог задатка, укључујући избор адекватних функција циља и мера растојања између дистрибуција.
- У четвртој фази ће, на основу претходних закључака, бити имплементиран алгоритамски радни оквир за импутацију, са подршком за рад над великим експресионим матрицама, уз истраживање могућности паралелизације и убрзања извршавања.
- Финална фаза обухвата темељну евалуацију предложеног решења на разноврсним скуповима података, поређење са постојећим приступима, као и анализу перформанси у погледу тачности, очувања биолошки значајних гена (маркер гени) и рачунарске ефикасности.

7. Закључак и предлог

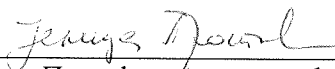
На основу приложене документације и стеченог целокупног утиска о теми докторске дисертације и досадашњег научног рада кандидата Лазара Смиљковића, мастер инжењера електротехнике и рачунарства, Комисија је дошла до следећих закључака:

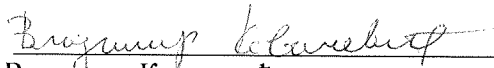
1. Кандидат испуњава све формалне и суштинске потребне услове прописане Законом о високом образовању, Статутом Универзитета у Београду и Статутом Електротехничког факултета Универзитета у Београду, за израду докторске дисертације.
2. Предложена тема „Развој и евалуација метода за анализу и унапређење квалитета података просторне транскриптомике“ јесте научно заснована, а резултати који ће проистећи из израде ове докторске дисертације представљаће значајан и оригиналан допринос међународној научној заједници. Тема припада ужој научној области Софтверско инжењерство.

На основу свега претходно изложеног, Комисија са задовољством предлаже Наставно-научном већу Електротехничког факултета у Београду да кандидату Лазару Смиљковићу одобри израду докторске дисертације под насловом „Развој и евалуација метода за анализу и унапређење квалитета података просторне транскриптомике“ (енг. *“Development and evaluation of methods for analyzing and improving the quality of spatial transcriptomics data”*) под руководством ментора др Марка Мишића, ванредног професора Електротехничког факултета Универзитета у Београду.

У Београду, 14.07.2025. године

ЧЛАНОВИ КОМИСИЈЕ


др Јелица Протић, редовни професор
Универзитет у Београду – Електротехнички факултет


др Владимир Ковачевић, научни сарадник
Истраживачко-развојни институт за вештачку
интелигенцију Србије


др Милица Јанковић, ванредни професор
Универзитет у Београду – Електротехнички факултет