

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

I ПОДАЦИ О КОМИСИЈИ		
1. Датум и орган који је именовао комисију: 21. фебруар 2024. године Наставно-научно веће Природно-математичког факултета, Универзитета у Новом Саду, одлука бр. 0603-466/21-18		
2. Састав комисије у складу са <i>Правилима докторских студија Универзитета у Новом Саду</i> :		
1. <u>Радовановић Милош</u> презиме и име	<u>Редовни професор</u> звање	<u>Рачунарске науке, 12.12.2021.</u> ужа научна област и датум избора
<u>Природно-математички факултет, УНС</u> установа у којој је запослен-а		<u>председник</u> функција у комисији
2. <u>Сливка Јелена</u> презиме и име	<u>Ванредни професор</u> звање	<u>Примењене рачунарске науке и информатика, 25.06.2020.</u> ужа научна област и датум избора
<u>Факултет техничких наука, УНС</u> установа у којој је запослен-а		<u>члан</u> функција у комисији
3. <u>Шикопарија Бранко</u> презиме и име	<u>Научни саветник</u> звање	<u>Биологија, 19.06.2023.</u> ужа научна област и датум избора
<u>Институт Биосенс, УНС</u> установа у којој је запослен-а		<u>члан</u> функција у комисији
4. <u>Тешендић Данијела</u> презиме и име	<u>Ванредни професор</u> Звање	<u>Информациони системи 17.11.2020.</u> ужа научна област и датум избора
<u>Природно-математички факултет, УНС</u> установа у којој је запослен-а		<u>ментор</u> функција у комисији
<p>Именована комисија истиче да је кандидат Предраг Матавуљ у спровођењу експеримената и изради свог докторског рада има и ментора (одлука Сената Универзитета у Новом Саду број: 04-29/4 од 26.01.2022) са Института БиоСенс – истраживачко развојног института за информационе технологије биосистема у којем је био запослен, што је и наведено у захвалници на почетку докторске дисертације. Др Сања Брдар, виши научни сарадник (датум избора: 28.02.2023) није формално могла бити именована у комисију за оцену и одбрану докторске дисертације пошто је на породилском одсуству али ју је комисија консултовала у процесу оцене докторске дисертације и израде овог извештаја.</p> <p>(енг.)</p> <p>The appointed committee emphasizes that in the process of conducting experiments and writing the dissertation the candidate Predrag Matavulj has had a second advisor (University of Novi Sad Senate decision no. 04-29/4 dated 26 January 2022) from the BioSense Institute – research and development institute for information technologies in biosystems where he was employed, which is stated in the acknowledgments at the beginning of the dissertation. Dr Sanja Brdar, senior research associate (date of appointment: 28 February 2023) could not be formally appointed into the committee for evaluation and defense of the doctoral dissertation due to maternity leave, but the committee consulted her in the process of evaluating the doctoral dissertation and writing this report.</p>		

<b>II ПОДАЦИ О КАНДИДАТУ</b>
<p>1. Име, име једног родитеља, презиме: Предраг, Жељка, Матавуљ</p> <p>2. Датум рођења, општина, држава: 30.10.1993., Вуковар, Хрватска</p> <p>3. Назив факултета, назив претходно завршеног нивоа студија и стечени стручни/академски назив: Природно-математички факултет, Универзитет у Новом Саду, мастер академске студије Примењена математика – наука о подацима, мастер математичар</p> <p>4. Година уписа на докторске студије и назив студијског програма докторских студија: 2018. година, Докторске академске студије Информатика</p>
<b>III НАСЛОВ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ: TITLE OF THE DOCTORAL DISSERTATION:</b>
<p>Мулти-модална архитектура заснована на машинском учењу за класификацију полена у реалном времену (енг.) Multi-modal architecture based on machine learning for real-time pollen classification</p>
<b>IV ПРЕГЛЕД ДОКТОРСКЕ ДИСЕРТАЦИЈЕ: OVERVIEW OF THE DOCTORAL DISSERTATION:</b>
<p>Докторска дисертација има следећи облик: 9 / 263 / 207 / 14 / 73 / 0 / 2 (број поглавља / страна / лит. цитата / табела / слика / графика / прилога)</p> <p>Докторска дисертација је написана на енглеском језику, са кључном документацијом писаном на српском и енглеском језику, као и кратким апстрактном писаним на српском и енглеском језику и продуженим апстрактном писаним на српском језику, латиничним писмом. Садржај докторске дисертације је распоређен у 9 издвојених поглавља:</p> <ol style="list-style-type: none"> <li>1. Увод (стр. 1–10): У уводном поглављу је дат кратак преглед теме истраживања у области, дефинисане су хипотезе докторског истраживања, као и кратак преглед организације по поглављима и научни допринос.</li> <li>2. Основе вештачке интелигенције и неуронских мрежа (стр. 11–24): Ово поглавље садржи опсежан преглед литературе како би се објаснили основни концепти вештачке интелигенције и неуронских мрежа.</li> <li>3. Методологија прикупљања и пред-обrade података (стр. 25–60): У трећем поглављу детаљно су представљени подаци и технологије коришћене у раду ове тезе. Обављено је темељно истраживање података, као и примена метода за елиминацију шума из података. Такође, описане су и технике пред-обrade података које су кључне за процес машинског учења.</li> <li>4. Експериментални приступи у детекцији честица суспендованих у ваздуху (стр. 61–113): У четвртном поглављу описана је методологија за класификацију полена суспендованог у ваздуху. Поглавље детаљно описује методе и технике које су усвојене за имплементацију и тренинг модела, укључујући различите архитектуре неуронских мрежа. Такође, поглавље испитује утицај шума на перформансе модела и представља дизајн и резултате ансамбл учења. Коначно, представљена је компаративна анализа између аутоматских класификација и мерења стандардном Хирстовом методом која подразумева идентификацију на основу анализе микроскопом и морфолошких карактеристика.</li> </ol>

5. Побољшање корелације модела применом трансфер учења (стр. 114–131): У петом поглављу истражује се развој методологије трансфер учења за искориштавање ручно аотираних мерења ради побољшања модела за аутоматску класификацију.
6. Адаптација домена за употребу модела на различитим уређајима (стр. 132–146): Ово поглавље фокусира се на стратегије адаптације домена које олакшавају пренос модела на различите уређаје са минималним смањењем перформанси.
7. Разоткривање механизма одлучивања модела (стр. 147–172): У седмом поглављу истражује се аспект објашњивости модела са нагласком на кумулативној градијентној методи. Поглавље дискутује о значајним карактеристикама и наученим обрасцима модела неуронских мрежа за класификацију полена.
8. Систем за аутоматску детекцију честица суспендованих у ваздуху у реалном времену (стр. 173–180): У осмом поглављу описује се софтверски систем који покрива процес од моделирања базе података за складиштење резултата класификације до дизајна кориснички оријентисаног интерфејса за веб сајт и апликацију.
9. Закључак и будући рад (стр. 181–185): У последњем поглављу дати су закључци истраживања ове дисертације, као и идеје за будући рад.

(енг.)

The doctoral dissertation has the following format:

9 / 263 / 207 / 14 / 73 / 0 / 2

(number of chapters, pages, literary citations, tables, figures, graphics, appendices)

The doctoral dissertation is written in English, with key documentary information written in Serbian and English, as well as a short abstract written in Serbian and English, and an extended abstract written in Serbian using the Latin script. The content of the doctoral dissertation is arranged in 9 separate chapters:

1. Introduction (pp. 1–10): The introductory chapter provides a brief overview of the topic and research in the field, defines the hypotheses of the doctoral research, as well as a brief overview of the organization by chapters and scientific contributions.
2. Foundations of Artificial Intelligence and Neural Networks (pp. 11–24): This chapter contains an extensive literature review to explain the basic concepts of artificial intelligence and neural networks.
3. Methodology for Data Collection and Preprocessing (pp. 25–60): In the third chapter, the data and technologies used in this thesis are presented in detail. A thorough investigation of the data was conducted, as well as the application of methods for removing noise from the data. Additionally, techniques for data preprocessing, which are crucial for the process of machine learning, are described.
4. Experimental Approaches in Airborne Particle Detection (pp. 61–113): The fourth chapter describes the methodology for the classification of air particles. The chapter details the methods and techniques adopted for the implementation and training of models, including various neural network architectures. It also examines the impact of noise on model performance and articulates the design and execution of ensemble learning techniques. Finally, a comparative analysis between automatic classifications and standard Hirst type pollen measurements that relies on microscopy and specific morphology features for classification is presented.
5. Enhancing Model Correlation with Transfer Learning (pp. 114–131): The fifth chapter explores the development of transfer learning methodology to leverage manually annotated measurements to improve models for automatic classification.
6. Domain Adaptation for Cross-Device Model Generalization (pp. 132–146): This chapter focuses on domain adaptation strategies that facilitate the transfer of models to different devices with minimal performance reduction.

7. Decoding the Model's Decision Mechanisms (pp. 147–172): The seventh chapter explores the aspect of model explainability with an emphasis on the cumulative gradient method. The chapter discusses significant features and the learned patterns of neural network models.
8. Real-Time System for Automatic Aerosol Detection (pp. 173–180): The eighth chapter describes a software system that covers the process from the database modeling for storing classification results to the design of a user-oriented interface for the website and application.
9. Conclusion and Future Work (pp. 181–185): The last chapter presents the conclusions of the research of this dissertation, as well as ideas for future work.

**V ВРЕДНОВАЊЕ ПОЈЕДИНИХ ДЕЛОВА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:  
EVALUATION OF INDIVIDUAL PARTS OF THE DOCTORAL DISSERTATION:**

**НАСЛОВ** докторске дисертације је јасно дефинисан и формулисан је у складу са садржајем студије. **Комисија констатује да је наслов дисертације разумљив и јасно дефинише садржај истраживања.**

У поглављу **Увод** се кратко описује предмет истраживања са прецизно дефинисаним циљевима докторске дисертације. Указано је на важност прецизне аутоматске класификације честица које су суспендоване у атмосфери. Дати су доприноси докторске дисертације, те је описана структура тезе.

У другом поглављу под називом **Основе вештачке интелигенције и неуронских мрежа** дате су дефиниције неуронских мрежа као и објашњење њиховог тренинга, што је неопходно за разумевање остатка тезе. У истом поглављу су додатно уведени појмови трансфер учења, адаптације домена, и објашњиве вештачке интелигенције.

Посебно значајно за ову дисертацију је што се бави аутоматизацијом класификације честица суспендованих у ваздуху кроз призму података, као што су подаци добијени помоћу уређаја Рапид-Е, који ласерима аутоматски обасјава појединачне честице, и снима податке расејане светлости, као и спектра и трајања флуоресценције. Такви подаци су захтевни за анализу и интерпретацију због своје комплексности па је за њихову употребу неопходно применити напредне технологије вештачке интелигенције што је у овом истраживању и урађено и представља посебан допринос науци и области. Детаљан опис коришћених технологија и података је дат у трећем поглављу под називом **Методологија прикупљања и пред-обrade података**. Додатно, ово поглавље истражује податке и налази методе уклањања шума из података.

**Комисија закључује да су поглавља Увод, Основе вештачке интелигенције и неуронских мрежа и Методологија прикупљања и пред-обrade података (прво, друго и треће поглавље) концептуално и садржајно одговарајућа за дефинисану тему истраживања. Поглавља садрже детаљно наведену релевантну литературу у истраживачкој области. Кроз прва три поглавља дат је опширан увод у истраживање, дефинисане су хипотезе, дат је преглед области и значајних теоријских концепата и технологије која је примењена у експерименталном делу истраживања.**

Поглавље **Експериментални приступи у детекцији честица суспендованих у ваздуху** описује методологију за класификацију честица суспендованих у ваздуху. Поглавље детаљно описује методе и технике које су усвојене за имплементацију и тренинг модела, укључујући различите архитектуре неуронских мрежа. Такође, испитује утицај шума на перформансе модела и представља дизајн и експерименте техника ансамбл учења. Коначно, представљена је компаративна анализа између аутоматских класификација и мерења стандардном Хирстовом методом која подразумева идентификацију на основу анализе микроскопом и морфолошких карактеристика. Један део резултата које описује ово поглавље је објављен у радовима [1, 3, 7, 8].

У поглављу **Побољшање корелације модела применом трансфер учења** истражује се развој методологије трансфер учења за искориштавање ручно аотираних мерења за побољшање модела за

аутоматску класификацију. Коришћени су модели из поглавља 4, који су сада до-тренирани са ручно аотираним подацима. Методологија овог поглавља је детаљно анализирана методом крос-валидације, те су добијени резултати показали значајно побољшање у односу на почетне моделе. Резултати овог поглавља су презентовани на међународној конференцији [12], те су послати у међународни часопис од кога се још чека одлука.

Поглавље **Адаптација домена за употребу модела на различитим уређајима** фокусира се на стратегије адаптације домена које олакшавају пренос модела на различите уређаје са минималним смањењем перформанси. Уочено је да модели који се истренирају на подацима са једног уређаја губе перформансе када се тестирају на подацима другог уређаја. Користећи методу обрнутог градијента, ово поглавље детаљно испитује пренос модела са једног на други уређај и показује значајно побољшање у односу на почетне моделе. Резултати овог поглавља су објављени у конференцијском раду [11] и у раду у међународном часопису [5].

У поглављу **Разоткривање механизма одлучивања модела** истражује се аспект објашњивости модела са нагласком на кумулативној градијентној методи. Неуронске мреже се често називају црним кутијама јер резултати нису објашњиви. Ово поглавље дискутује о значајним карактеристикама и наученим обрасцима модела неуронских мрежа за класификацију полена. Резултати овог истраживања показали су да модел може да учи и на шуму, а објављени су у међународном часопису [2].

Последње поглавље пре самог закључка, под називом **Систем за аутоматску детекцију честица суспендованих у ваздуху у реалном времену**, описује софтверски систем који покрива процес од имплементације базе података за складиштење резултата класификације, до дизајна кориснички оријентисаног интерфејса за веб сајт и апликацију. Цео систем представља практичну примену резултата ове докторске дисертације која може помоћи људима алергичним на полен. Резултати овог истраживања објављени су у међународном часопису [4].

**Комисија оцењује да су поглавља Експериментални приступи у детекцији честица суспендованих у ваздуху, Побољшање корелације модела применом трансфер учења, Адаптација домена за употребу модела на различитим уређајима, Разоткривање механизма одлучивања модела, и Систем за аутоматску детекцију честица суспендованих у ваздуху у реалном времену (четврто, пето, шесто, седмо и осмо поглавље) јасно и прегледно написана, да је структура текста добро конципирана и да су резултати истраживања оригинални, примењиви у пракси и релевантни у области. Кандидат Предраг Матавуљ је студиозно и објективно, користећи напредне технологије машинског учења и вештачке интелигенције, детаљно анализирао податке и интерпретирао резултате у одговарајућем контексту.**

У поглављу **Закључак и будући рад** су представљени закључци из целокупне студије која је изведена у току докторског истраживања. Потврђене су хипотезе које су дефинисане у уводном делу тезе и продискутовани су главни закључци истраживања. У докторској дисертацији представљен је детаљан приступ аутоматизације класификације полена суспендованог у ваздуху, коришћене су напредне методе за класификацију, трансфер учење, адаптацију домена, које су значајно побољшале резултат класификације. Резултати тезе су практично представљени на мобилној и веб апликацији.

**Комисија закључује да су остварени планирани циљеви истраживања у оквиру докторске дисертације, да су постигнути значајни научни резултати у области, као и да је закључак рада јасно написан.**

Након поглавља **Закључак** следи цитирана литература. У оквиру целе дисертације наведено је 207 извора.

**Комисија сматра да је литература цитирана на одговарајући начин, да је избор литературе добар и да наведене референце представљају значајне научне доприносе у области.**

(EHR.)

The **TITLE** of the doctoral dissertation is clearly defined and formulated in accordance with the content of the study.

**The committee notes that the title of the dissertation is understandable and clearly defines the content of the research.**

In the **Introduction** chapter, the subject of research is briefly described with precisely defined objectives of the doctoral dissertation. This chapter highlights the importance of precise automatic classification of air particles. The contributions of the doctoral dissertation are given, and the structure of the thesis itself is described.

In the second chapter titled **Foundations of Artificial Intelligence and Neural Networks**, definitions of neural networks are given as well as an explanation of their training, necessary for understanding the rest of the thesis. In the same chapter, concepts of transfer learning, domain adaptation, and explainable artificial intelligence are additionally introduced.

Particularly significant for this dissertation is that it deals with the automation of air particle classification through the prism of data, such as data obtained using the Rapid-E device, which automatically records particles with lasers, where data on scattered light, as well as fluorescence spectrum and lifetime are obtained. Such data are challenging for analysis and interpretation due to their complexity, so the application of advanced artificial intelligence technologies is necessary for their use, which has been done in this research and represents a special contribution to science and the field. A detailed description of the used technologies and data is given in the third chapter titled **Methodology for Data Collection and Preprocessing**. Additionally, this chapter explores data and finds methods for removing noise from the data.

**The committee concludes that the chapters Introduction, Foundations of Artificial Intelligence and Neural Networks, and Methodology for Data Collection and Preprocessing (first, second, and third chapters) are conceptually and content-appropriate for the defined research topic. The chapters contain a detailed list of relevant literature in the research area. Through the first three chapters, an extensive introduction to the research is given, hypotheses are defined, a literature review and significant theoretical concepts of technology applied in the experimental part of the research are presented.**

The chapter **Experimental Approaches in Airborne Particle Detection** describes the methodology for classifying air particles. The chapter details the methods and techniques adopted for the implementation and training of models, including various neural network architectures. It also examines the impact of noise on model performance and articulates the design and execution of ensemble learning techniques. Finally, a comparative analysis between automatic classifications and standard Hirst type pollen measurements that relies on microscopy and specific morphology features for classification is presented. Part of the results described in this chapter has been published in the works [1, 3, 7, 8].

In the chapter **Enhancing Model Correlation with Transfer Learning**, the development of transfer learning methodology is explored to leverage manually annotated measurements to improve models for automatic classification. Models from chapter 4, now retrained with new, manually annotated data, were used. The methodology of this chapter is thoroughly analyzed with cross-validation, and the results have shown significant improvement over the initial models. The results of this chapter were presented at an international conference [12] and submitted to an international journal awaiting a decision.

The chapter **Domain Adaptation for Cross-Device Model Generalization** focuses on domain adaptation strategies that facilitate the transfer of models to different devices with minimal performance reduction. It was observed that models trained on data from one device lose performance when tested on data from another device. Using the reverse gradient method, this chapter thoroughly examines the transfer of models

from one device to another and shows significant improvement over the initial models. The results of this chapter have been published in the conference paper [11] and in an article in an international journal [5].

In the chapter **Decoding the Model's Decision Mechanisms**, the aspect of model explainability is explored with an emphasis on the cumulative gradient method. Neural networks are often called black boxes because their results are not explainable. This chapter discusses significant features and learned patterns of neural network models for pollen classification. The results of this research showed that the model can learn even from noise, and they have been published in an international journal [2].

The penultimate chapter before the conclusion, titled **Real-Time System for Automatic Aerosol Detection**, describes a software system that covers the process from the implementation of a database for storing classification results to the design of a user-oriented interface for the website and application. The entire system represents the practical application of the results of this doctoral dissertation, which can help people allergic to pollen. The results of this research have been published in an international journal [4].

**The committee concludes that the chapters Experimental Approaches in Airborne Particle Detection, Enhancing Model Correlation with Transfer Learning, Domain Adaptation for Cross-Device Model Generalization, Decoding the Model's Decision Mechanisms and Real-Time System for Automatic Aerosol Detection (fourth, fifth, sixth, seventh, and eighth chapters) are clearly and comprehensively written, that the text structure is well-conceived, and that the research results are original, applicable in practice, and relevant in the field. The candidate Predrag Matavulj has objectively and thoroughly analyzed data using advanced machine learning and artificial intelligence technologies, and interpreted results in the appropriate context.**

In the chapter Conclusion and Future Work, conclusions from the entire study conducted during the doctoral research are presented. The hypotheses defined in the introductory part of the thesis are confirmed, and the main conclusions related to the research are discussed. The doctoral dissertation presents a detailed approach to the automation of air particle classification, specifically pollen, using advanced methods for classification, transfer learning, domain adaptation, which have significantly improved classification results. The results of the thesis are practically presented on the website and mobile application.

**The committee concludes that the planned research objectives within the doctoral dissertation have been achieved, that significant scientific results have been achieved in the field, and that the conclusion of the work is clearly written.**

After the Conclusion chapter follows cited literature. Within the entire dissertation, 207 resources are cited.

**The committee believes that the literature is cited appropriately, that the choice of literature is good, and that the listed references represent significant scientific contributions in the field.**

**VI СПИСАК НАУЧНИХ И СТРУЧНИХ РАДОВА КОЈИ СУ ОБЈАВЉЕНИ ИЛИ ПРИХВАЋЕНИ ЗА ОБЈАВЉИВАЊЕ НА ОСНОВУ РЕЗУЛТАТА ИСТРАЖИВАЊА У ОКВИРУ РАДА НА ДОКТОРСКОЈ ДИСЕРТАЦИЈИ:**

**LIST OF PUBLISHED OR ACCEPTED FOR PUBLISHING PAPERS THAT ARE RESULT OF THE DOCTORAL DISSERTATION:**

Истраживање спроведено у овој докторској дисертацији прошло је кроз строг процес рецензије, резултирајући објављивањем следећих оригиналних радова у међународним часописима:  
(енг.)

The research conducted in this doctoral thesis underwent a rigorous peer review process, resulting in the following original papers published in international journals:

1. Matavulj, P., Panić, M., Sikoparija, B., Tesendic, D., Radovanovic, M., Brdar, S. 2023. Advanced CNN Architectures for Pollen Classification: Design and Comprehensive Evaluation, Applied Artificial Intelligence, 37(1), 2157593, doi: 10.1080/08839514.2022.2157593 (M22)

2. Brdar, S., Panic, M., Matavulj, P., Stankovic, M., Bartolic, D. Šikoparija, B. 2023. Explainable AI for unveiling deep learning pollen classification model based on fusion of scattered light patterns and fluorescence spectroscopy. *Scientific Reports* 13, 3205. doi: 10.1038/s41598-023-30064-6 (M21)
3. Šauliene, I., Šukiene, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., Brdar, S., Panić, M., Šikoparija, B., Clot, B., Crouzy, B., Sofiev, M. 2019. Automatic pollen recognition with the Rapid-E particle counter: the first-level procedure, experience and next steps. *Atmospheric Measurement Techniques*, 12, 3435–3452. doi: 10.5194/amt-12-3435-2019 (M21)
4. Tesendić, D., Krsticev, D. B., Matavulj, P., Brdar, S., Panić, M., Minić, V., and Šikoparija, B. 2022. RealForAll: real-time system for automatic detection of airborne pollen. *Enterprise Information Systems*, 16(5), 1793391. doi: 10.1080/17517575.2020.1793391 (M21)
5. Matavulj, P., Cristofori, A., Cristofolini, F., Gottardini, E., Brdar, S., Šikoparija, B. 2022. Integration of reference data from different Rapid-E devices supports automatic pollen detection in more locations. *Science of the Total Environment*, 851(2), 158234. doi: 10.1016/j.scitotenv.2022.158234 (M21a)
6. Smith, M., Matavulj, P., Mimić, G., Panić, M., Grewling, L., Šikoparija, B. 2022. Why should we care about high temporal resolution monitoring of bioaerosols in ambient air? *Science of The Total Environment*, 826, 154231. doi: 10.1016/j.scitotenv.2022.154231 (M21a)
7. Šikoparija, B., Matavulj, P., Mimić, G., Smith, M., Grewling, L., Podrašćanin, Z. 2022. Real-time automatic detection of starch particles in ambient air. *Agricultural and Forest Meteorology*, 323, 109034. doi: 10.1016/j.agrformet.2022.109034 (M21a)
8. Maya-Manzano, J. M., et al. 2023. Towards European automatic bioaerosol monitoring: Comparison of 9 automatic pollen observational instruments with classic Hirst-type traps. *Science of The Total Environment*, 866, 161220, doi: 10.1016/j.scitotenv.2022.161220 (M21a)
9. Šikoparija, B., Mimić, G., Matavulj, P., Panić, M., Simović, I., Brdar, S. 2019. Short communication: Do we need continuous sampling to capture variability of hourly pollen concentrations? *Aerobiologia*, 36, 3–7. doi: 10.1007/s10453-019-09575-1 (M22)
10. Simović, I., Matavulj, P., Šikoparija, B. 2023. Manual and automatic quantification of airborne fungal spores during wheat harvest period. *Aerobiologia* 39, 227–239, doi: 10.1007/s10453-023-09788-5 (M22)

Кључни нови концепти представљени су на познатим међународним конференцијама како би се олакшала дискусија о експериментима и добијеним резултатима са заједницама из области рачунарских наука и науке о подацима:

(енг.)

The critical novel concepts were presented at well-known international conferences to facilitate discussions of the experiments and the results obtained with the computer science and data science communities:

11. Matavulj, P., Brdar, S., Racković, M., Šikoparija, B., Athanasiadis, I. N. 2021. Domain adaptation with unlabeled data for model transferability between airborne particle identifiers. In *Proceedings of the 17th International Conference on Machine Learning and Data Mining*, 147–158, July 19, 2021, New York, USA (M33)
12. Matavulj, P., Jelic, S., Brdar, S., Radovanovic, M., Tesendic, D., Šikoparija, B. 2023. Reoptimizing neural networks for pollen classification, *The International Conference on Optimization and Learning*, May 3–5, 2023, Malaga, Spain (poster presentation)



Увидом у публиковане научне радове, као и саопштења са скупова, у којима су објављени делови резултата докторске дисертације, Комисија констатује да кандидат испуњава услове прописане за одбрану докторске дисертације.

(енг.)

**Upon reviewing the published scientific papers and conference proceedings in which parts of the results of the doctoral dissertation are published, the Committee concludes that the candidate meets the conditions prescribed for the defense of the doctoral dissertation.**

## **VII ЗАКЉУЧЦИ ОДНОСНО РЕЗУЛТАТИ ИСТРАЖИВАЊА: CONCLUSIONS AND RESULTS OF THE RESEARCH:**

На основу резултата истраживања, који су постигнути реализацијом постављених циљева и задатака уз строгу примену научне методологије и коришћењем обимне мултидисциплинарне научне литературе у домену класификације честица суспендованих у ваздуху, машинског учења, неуронских мрежа, алгоритама и рачунарских технологија, изведени су закључци који доказују иницијално постављене хипотезе. У наставку ћемо детаљније анализирати сваку од хипотеза и објаснити методе којима су оне потврђене.

- Истраживање је обухватало развој робусних метода за прикупљање и обраду података, који су били од суштинског значаја за анализу разлика у подацима и карактеристика шума. Забележене су значајне разлике у подацима добијеним са различитих уређаја (Осијек и Нови Сад), које су укључивале варијације у осетљивости мерења или калибрацији. Развијена је стратегија за управљање шумом, нарочито за податке о спектру флуоресценције, где су касније временске ознаке класификоване као шум.
- Студија је такође усмерена на истраживање архитектура неуронских мрежа за класификацију 24 класе полена и две друге класе биоаеросола. Резидуалне мреже показале су се као најефикасније за једномодалну класификацију, достигнувши F1 резултат око 0,5. У вишемодалним подешавањима, једноставнији модели као што су ванила CNN мреже су пожељни због мањих рачунарских захтева и боље интерпретабилности. Најбоља вишемодална мрежа је постигла F1 резултат од 0,69 за 26 класа.
- Даље, перформансе аутоматизованог система су упоређене са мерењима стандардном Хирстовом методом која подразумева идентификацију полена на основу анализе микроскопом и морфолошких карактеристика, откривајући умерене до јаке корелације у одређеним класама полена. На пример, 8 од 23 класа полена показало је корелације веће од 0,65, а 14 класа показало је корелације изнад 0,5 на просеку током година.
- Испитана је отпорност мрежа на шум у подацима, остварујући ефикасност од око 42–48% чак и са компромитованим изворима података. Највећа отпорност је уочена када је податак трајања флуоресценције компромитован, што указује на то да се модел више ослања на одлуке базиране на спектру флуоресценције и сликама расејане светлости.
- Применом трансфер учења дошло је до значајног унапређења у аутоматској идентификацији, са 23 од 29 случајева који су показали статистички значајно побољшање. Модели су у просеку показали повећање корелације за 22,52% и значајно смањење стандардне девијације за 38,05%.
- Такође су истражене технике адаптације домена како би се адресирао проблем пада перформанси (13–26%) када су модели обучени на једном уређају, а тестирани на другом, постигавши побољшање од 11% са овом методом.
- Технике објашњиве вештачке интелигенције су учиниле одлуке модела неуронских мрежа транспарентним, идентификујући значајан утицај спектра флуоресценције, нарочито између 364–538 нанометара.
- Теза је комплетирана приказом система RealForAll, софтверске апликације за презентацију концентрације честица полена у ваздуху у реалном времену, која премодулира јаз између истраживања и примењене информатике.

**Комисија позитивно оцењује дефинисане хипотезе, резултате и закључке истраживања.**

(eHR.)

Based on the results of the research, during which the set objectives and tasks were realized using scientific methodology and utilizing vast scientific literature from multidisciplinary research related to the classification of airborne particles, machine learning, neural networks, algorithms, and computer technologies, conclusions were drawn that confirm the initially defined hypotheses. We will briefly review the specific hypotheses and explain how they were confirmed.

- The research involved developing robust methodologies for data collection and preprocessing, crucial for handling data discrepancies and noise. Notable differences in data from different devices (Osijek and Novi Sad) were observed, with variations in measurement sensitivity or calibration. A strategy was developed for managing noise, particularly in fluorescence spectrum data, with later timestamps classified as noise.
- The study explored neural network architectures for classifying 24 pollen classes and two other bioaerosol classes. Residual networks were most effective for unimodal classification, achieving around a 0.5 F1 score. In multimodal settings, simpler models like vanilla CNNs were practical due to lower computational demands and better interpretability. The best multimodal network achieved an F1 score of 0.69 across 26 classes.
- The real-world applicability of the models was emphasized, with the necessity to test them in unpredictable environments. The automatic system's performance was compared with standard Hirst type pollen measurements that rely on microscopy and specific morphology features for classification, revealing moderate to strong correlations in certain pollen classes. For example, 8 out of 23 pollen classes showed correlations greater than 0.65, and 14 classes showed correlations above 0.5 when averaged over the years.
- The networks' resilience to noise was tested, maintaining about 42–48% effectiveness even with compromised data sources. The highest resilience was observed when the lifetime feature was corrupted, indicating the model's reliance on fluorescence spectrum and scattered light image data.
- Transfer learning was employed to enhance automatic identification, leading to an average 22.52% increase in correlation and a 38.05% reduction in standard deviation across 29 evaluated cases, with 23 showing statistically significant improvements.
- Domain adaptation techniques were explored to address the performance decline (13–26%) when models trained on one device were tested on another, achieving an 11% improvement with these techniques.
- Explainable AI techniques helped make AI decision-making transparent, with significant influence identified in the spectrum data modality, especially between 364–538 nm.
- The thesis concluded with the introduction of the RealForAll system, a software application for real-time airborne particle concentration presentation, bridging the gap between research and applied informatics.

**The Committee positively evaluates the defined hypotheses, results, and conclusions of the research.**

**VIII ОЦЕНА НАЧИНА ПРИКАЗА И ТУМАЧЕЊА РЕЗУЛТАТА ИСТРАЖИВАЊА:  
ASSESSMENT OF THE PRESENTATION AND EXPLANATION OF THE RESULTS:**

Кандидат Предраг Матавуљ је на јасан и одговарајући начин изложио, протумачио и представио резултате спроведеног истраживања кроз детаљан приказ примене теоријских концепата и експерименталних резултата. На основу резултата су изведени закључци који одговарају предмету студије и пружају одговоре на постављене циљеве истраживања. Структура рада поклапа се са планом и дефинисаним циљевима у пријави теме. Наведена литература коришћена у раду је савремена, обимна и релевантна за тему истраживања. Резултати су научно оправдани и подржани бројним изворима из литературе. На основу начина приказивања и тумачења података, може се констатовати да рад садржи оригиналне научне резултате који задовољавају захтеве нивоа докторске дисертације.

(енг.)

The PhD candidate Predrag Matavulj presented, explained, and discussed the results of the research in a clear and appropriate manner regarding theory and practical experiments. Based on the results, conclusions were drawn that correspond to the subject of the study and provide answers to the research objectives. The structure of the paper matches the plan and defined objectives. The cited literature used in the paper is modern, extensive, and relevant to the topic of the research. The results are scientifically justified and supported by numerous literature sources. Based on the way the data is presented and interpreted, it can be concluded that the research work contains original scientific results that meet the requirements of the doctoral dissertation level.

**IX КОНАЧНА ОЦЕНА ДОКТОРСКЕ ДИСЕРТАЦИЈЕ:  
FINAL NOTE ABOUT DOCTORAL DISSERTATION:**

1. Да ли је дисертација написана у складу са образложењем наведеним у пријави теме?

**Комисија оцењује да је докторска дисертација написана у складу са образложењем наведеним у пријави теме.**

(енг.)

1. Is the dissertation written in accordance with the explanation provided in the thesis proposal?

**The committee assesses that the doctoral dissertation is written in accordance with the explanation provided in the thesis proposal.**

2. Да ли дисертација садржи све битне елементе?

**Комисија оцењује да дисертација садржи све битне елементе оригиналног научног рада: јасно дефинисану тему истраживања, одговарајућу методологију уз употребу напредних технологија машинског учења и вештачке интелигенције, преглед постојеће литературе у мултидисциплинарним областима које обухватају аеробиологију, тачније класификацију полена, рачунарске науке, посебне области у науци о подацима и вештачкој интелигенцији. Такође, резултати су представљени на јасан начин и њихово значење и значај су продискутовани у складу са захтевима научног рада.**

(енг.)

2. Does the dissertation contain all the essential elements?

**The committee assesses that the dissertation contains all the essential elements of original scientific work: a clearly defined research topic, appropriate methodology using advanced machine learning and artificial intelligence technologies, a review of existing literature in multidisciplinary fields that encompass aerobiology, specifically pollen classification, computer science, special areas in data science, and artificial intelligence. Additionally, the results are presented in a clear manner, and their meaning and significance are discussed in accordance with the requirements of scientific work.**

3. По чему је дисертација оригиналан допринос науци?

Ово докторско истраживање помаже у развијању области машинског учења и информатике кроз нове доприносе у пет кључних области. Прво, оптимизује моделе машинског учења истраживањем различитих архитектура неуронских мрежа и стратегија класификације честица суспендованих у ваздуху. Ово истраживање се бави изазовима шума у подацима и њиховој варијабилности, посебно

када се они разликују од стандардних података коришћених за евалуацију мрежа. Друго, истраживање унапређује класификацију полена техникама трансфер учења дотрениравањем постојећих модела. Треће, унапређује се класификација на различитим уређајима техникама адаптације домена, са фокусом на ублажавање утицаја специфичног шума уређаја како би се повећала прецизност модела. Четврто, теза помаже напретку у области објашњиве вештачке интелигенције проширењем обима кумулативних градијентних метода за побољшану транспарентност и интерпретативност модела. Коначно, доприноси примењеној информатици дати су кроз развој софтверског система са приступачном базом података за презентацију концентрације полена у реалном времену, што представља примену истраживања. Колективно, ови доприноси не само да показују ефикасност уређаја Rapid-E у класификацији полена, већ и истичу значај квалитета података, као и разноликости података у развоју робусних рачунарских модела. Научна открића представљена у овој тези имају потенцијал да значајно унапреде оперативну ефикасност уређаја Rapid-E, што доприноси напретку у мониторингу алергена. Ово истраживање уводи тачније, поузданије и широко доступније методе за детекцију и анализу биоаеросола. Релевантност и значај ових доприноса поткрепљени су објављивањем бројних радова у међународним научним часописима и зборницима конференција.

#### **Напомена**

Докторска дисертација је у библиотеци Природно-математичког факултета прошла проверу оригиналности применом софтвера iThenticate за детекцију сличности који је показао да „индекс сличности“ (енг. Similarity Index) износи 17%, од којих је бар 14% преклапања текста везано за радове самог кандидата. Применом софтвера утврђено је да од 117 извора са којима је утврђено преклапање, у 95% извора је пронађено мање од 1% подударности. Извори у којима је пронађено 1% или више подударности (извори 1, 2, 3, 4, 5, и 6 у извештају сличности) представљају научне радове где је кандидат Предраг Матавуљ први аутор или један од аутора. Остатак преклапања од мање од 3% односи се на опште појмове и често коришћене фразе, највише из прегледног поглавља Основе вештачке интелигенције и неуронских мрежа.

**Комисија оцењује да је дисертација резултат оригиналног научног рада, и да садржи научно значајне податке.**

(енг.)

3. What makes the dissertation an original contribution to science?

This doctoral research significantly advances the fields of machine learning and informatics by offering novel contributions in five key areas. Firstly, it optimizes machine learning models through an exploration of various neural network architectures and classification strategies for airborne particle classification. This work addresses the challenges of data noise and variability, particularly in datasets differing from standard benchmarks. Secondly, the research enhances transfer learning techniques by improving existing models and pioneering hybrid learning approaches. Thirdly, improvements in domain adaptation techniques are given, focusing on mitigating device-specific noise impacts to increase model precision across different devices. Fourthly, the thesis advances explainable AI (xAI) by expanding the scope of cumulative gradient methods for enhanced model transparency and interpretability. Lastly, it contributes to practical informatics through the development of a user-friendly software system with an accessible database for real-time airborne particle concentration presentation, blending rigorous research with application. Collectively, these contributions not only shed light on the efficacy of the Rapid-E device in airborne pollen detection but also emphasize the importance of data quality, diversity, and representation in developing robust computational models. The findings are poised to significantly enhance the operational efficiency of Rapid-E, thereby advancing environmental allergen surveillance and establishing more accurate, reliable, and accessible methods in bioaerosol detection and analysis. These contributions are confirmed by numerous publications in international journals and conference proceedings.

#### **Note:**

The doctoral dissertation underwent an originality check using the iThenticate software for similarity detection at the library of the Faculty of Sciences, which showed that the "Similarity Index" is 17%, of which at least 14% of the text overlap is related to the works of the candidate himself. The application of

the software determined that out of 117 sources with which overlap was identified, in 95% of the sources less than 1% similarity was found. The sources in which 1% or more similarity was found (sources 1, 2, 3, 4, 5, and 6 in the similarity report) represent scientific papers where the candidate Predrag Matavulj is the first author or one of the authors. The remaining overlap of less than 3% refers to general terms and frequently used phrases, most notably from the overview chapter of the Foundations of Artificial Intelligence and Neural Networks.

**The committee assesses that the dissertation is the result of original scientific work, and that it contains scientifically significant data.**

4. Који су недостаци дисертације и какав је њихов утицај на резултат истраживања?

**Комисија сматра да докторска дисертација нема значајних недостатака.**

(енг.)

4. What are the flaws of the dissertation and how do they affect the results of the research?

**The committee concludes that the doctoral dissertation does not have significant shortcomings.**

#### **X ПРЕДЛОГ (THE PROPOSAL):**

На основу наведеног, комисија предлаже:

Based on all stated facts, committee proposes:

**а) да се докторска дисертација прихвати, а кандидату одобри одбрана;  
to accept the doctoral dissertation and allow candidate to proceed with its defense;**

б) да се докторска дисертација врати кандидату на дораду (да се допуни односно измени);

в) да се докторска дисертација одбије.

Место и датум:

Нови Сад, 28.2.2024.

1. Милош Радовановић, редовни професор

\_\_\_\_\_, председник

2. Јелена Сливка, ванредни професор

\_\_\_\_\_, члан

3. Бранко Шикопарија, научни саветник

\_\_\_\_\_, члан

4. Данијела Тешендић, ванредни професор

\_\_\_\_\_, ментор

**НАПОМЕНА:** Члан комисије који не жели да потпише извештај јер се не слаже са мишљењем већине чланова комисије, дужан је да унесе у извештај образложење односно разлоге због којих не жели да потпише извештај и да исти потпише.