CERTIFICATE Of PARTICIPATION

Erkan Caner Ozkat

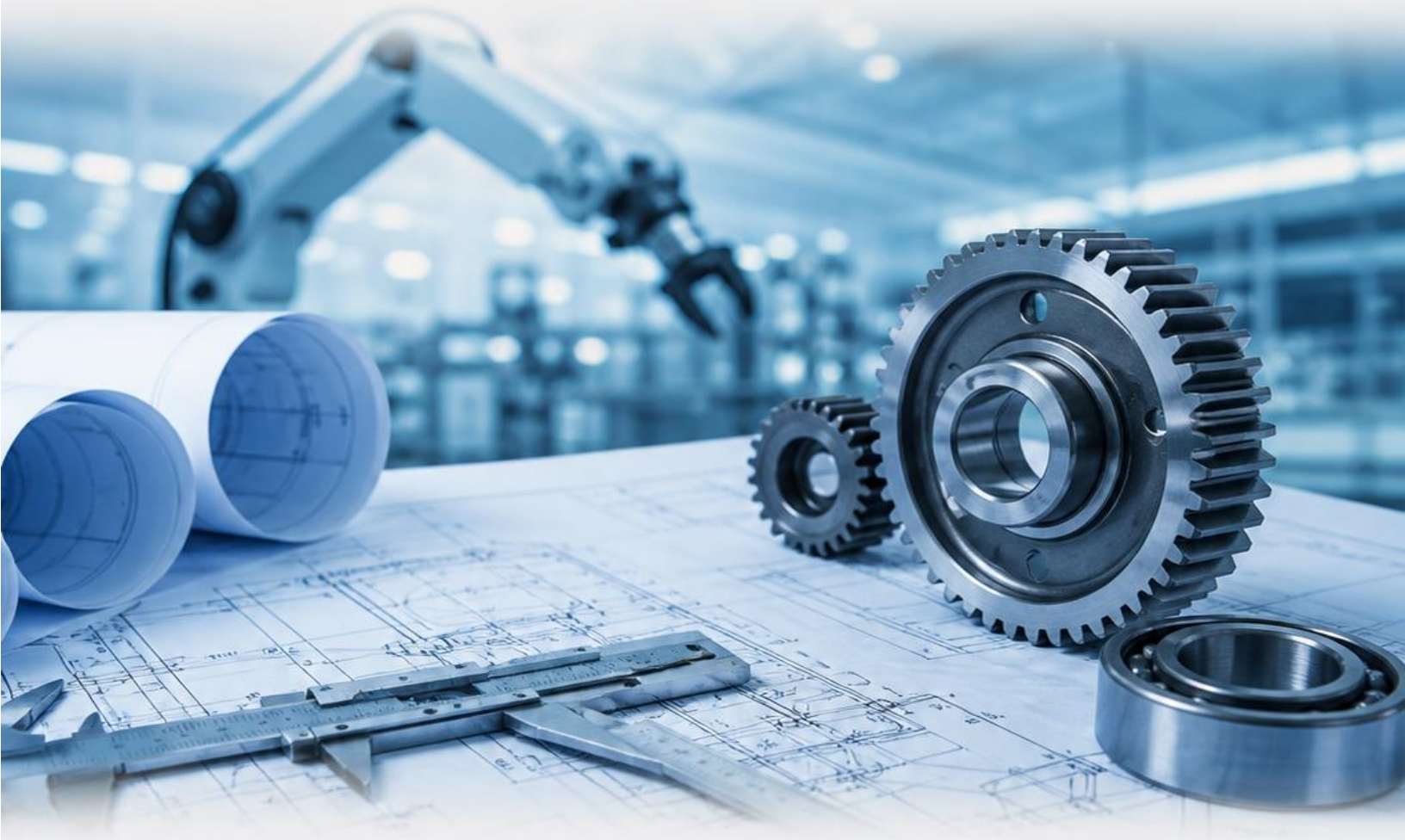
has participated in 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026 on August 25-26 in 2026 in Konya, Turkey.

PAPER TITLE	*Sequence Homology Inflates Reported Accuracy in Antimicrobial Peptide Prediction: A Machine Learning and Deep Learning Benchmark*
PRESENTATION TYPE	*Oral*

ICRIRET

All Sciences Academy

ALL SCIENCES ACADEMY



Proceeding Book of 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026

August 25-26, 2026: Konya, Turkey

PROCEEDING BOOK OF 5TH INTERNATIONAL CONFERENCE ON RECENT AND INNOVATIVE RESULTS IN ENGINEERING AND TECHNOLOGY ICRIRET 2026

PROCEEDING BOOK OF 5TH ICRIRET 2026:

25-26 August 2026

Konya, Turkey

Publication date: 05.09.2026

Publisher: All Sciences Academy

Proceeding Book of 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026

https://as-proceeding.com/index.php/icriret/home

5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026: edited by All Sciences Academy.

ISBN: 978-625-8839-81-4

As this is a serial book published, it has ISBN. All abstracts were peer reviewed.

ISBN: 978-625-8839-81-4

Printed in the Turkey with ISBN 978-625-8839-81-4

International Conference Website

https://as-proceeding.com/index.php/icriret/home

INTRODUCTION

We had the great honor of 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026. It was truly a great pleasure for us to greet a lot of participants from many different countries to attend ICRIRET 2026! We firmly believe that the conference will become an important international event in the field of cross-industry discussion about innovations in Academic Studies.

Three cooperating organizations supported the two-day conference. There were 127 papers accepted for presentation at ICRIRET 2026, contributed from different countries. We had plenary speeches and several well-known scientists and experts, to give invited talks at differentsessions.

The purpose of ICRIRET 2026 was to provide a forum for the participants to report and review innovative ideas, with up-to-date progress and developments, and discuss novel approaches to the application in the field of their own research areas and discuss challenges of doing science.

We sincerely hope that the exchange of ideas on doing research, science and improving education will help the participants, and international cooperation sharing the common interest will be enhanced.

On behalf the Organization Committee of ICRIRET 2026, we would like to heartily thank our cooperating organizations for all they have done for the conference. We would also like to thank the authors for theircontribution to the proceedings; the participants and friends of ICRIRET 2026, for their interest and efforts in helping us to make the conference possible; and the Editorial boards for their effective work and valuable advice, especially the ICRIRET 2026 secretariat and the ICRIRET 2026 staff, for their tireless efforts and outstanding services in preparing the conference and publishing the Proceedings.

Publisher: All Sciences Academy

TABLE OF CONTENTS

Introduction II
Table of Contents III
Conference Chairs X
Scientific Committees XI
Document of Academic Incentive XII
Document of International Conference XIII

Table List of Content

Content of Abstract Papers

Page No	Title	Authors	Country	Submission ID
1	Thermodynamic regression of a ternary system for extracting acetic acid from aqueous solutions using methyl tert-butyl ether using NRTL with PSO	Sabrina Hebboul, Oussama Bacha, Amel Hebboul	Algeria, Algeria, Algeria	Submission 376
2	CHRONOAMPEROMETRY STUDY OF SCALE INHIBITION IN HAMMA GROUNDWATER USING A QUERCETIN–ALGINATE MIXTURE.	Manel Boumagoura, Samira Ghizellaoui	Algeria, Algeria	Submission 377
3	Synergistic Removal of Salol via Iron Complexes under UVA Irradiation in Aqueous Solutions	Chaima BOUAZIZ, Nassira SERAGHNI, Nadra DEBBACHE, Tahar SEHILI	Algeria, Algeria, Algeria, Algeria	Submission 378
4	Ternary Oxide Heterojunction Engineering in MgO-CuO-ZnO Nanocomposites: Synergistic Charge Carrier Dynamics, Band Structure Modulation, and Multifunctional Optoelectronic Performance	Abdelbaki Obeidi, Thameur Obeidi, Mohamed Zitouni	Algeria, Algeria, Algeria	Submission 379
5	Micromechanical modeling of the effect of the granular skeleton on instantaneous and long-term intrinsic deformations of mortar	MOHAMMEDI Mohammed Nourelislam	Algeria	Submission 382
6	Spray Drying Technology Applied to Dairy Industrial Rinsing Waters: Pilot-Scale Atomization Using MP314 and Characterization of Reconstituted Powder Substrates	DERDICHE Abdelhamid, ZAMOUCHE ZERDAZI Rania	Algeria, Algeria	Submission 383
7	Physicochemical Characterization of Honey: Evaluation of Moisture Content and Refractive Index as Quality Indicators.	BELKHIRI Widad, DJAMAA Fatma	Algeria, Algeria	Submission 388
8	Secure Industrial IoT Networks: Hybrid AI–Blockchain Framework	Rimon Chandra Bhowmik, Jahid Hassan Sowrov, Md. Kausar Ahmed, Abu Nasim	Bangladesh, Bangladesh, Bangladesh, Bangladesh	Submission 392
9	Comparative Cradle-to-Gate Life Cycle Assessment of Burnt Clay and Fly Ash Bricks in Pakistan	Muhammad Hamza Bashir, Muhammad Farhan Javed, Saad Bilal Tabassum, Waqar Shafique	U.K., Pakistan, Germany, Pakistan	Submission 396
10	Parameter Estimation of the Rayleigh–Generalized Gamma Mixture Model Using Moment-Based Methods and Artificial Neural Networks	BENTOUMI AHMED, BENTOUMI MILOUD, BENDIB AHMED	Algeria, Algeria, Algeria	Submission 399
11	Closed-Form Representations for First-Kind Fredholm Integral Equations with Separable Kernels via Exponential Fourier Series and Lavrentiev-Type Regularisation	Ezgi Bal, Tahir Coşgun	Turkey, Turkey	Submission 400
12	Comparative Seismic Performance of High-Rise RC Buildings with and without Fluid Viscous Dampers	Shahzad Ali, Fahad Shahid Kawajah, Waqar Shafique, Muhammad Saad Duraiz	Cyprus, Pakistan, Pakistan, Cyprus	Submission 401
13	The Impact of Verbal Praise on Preschool Classroom Behavior	Afshan Naseem, Iqra Tariq	Pakistan, Pakistan	Submission 402
14	Valorization of Local Algerian Materials (Saharan Sand and Agro-Food Biochar) to Replace the Imported Sand/Anthracite Dual-Media for Seawater pretreatment: Case study of the “Myah-Tipaza” Desalination Plant	Rafik El Arslene DRA, Mounya ZENASNI, Malika MEDJAHDI, Nadia RAMDANI	Algeria, Algeria, Algeria, Algeria	Submission 405
15	Influence by Porosity for the Functionally Graded Plates Solar Used the Higher-Order Theory	Merdaci Slimane, Hadj Mostefa Adda, Merazi Mohamed, Bouchafa Ali	Algeria, Algeria, Algeria, Algeria	Submission 406
16	Automated MIG/MAG Process Parameter Calculation within a Computer-Aided Welding (CAW) Environment	Mohamed Rahou, Fethi SEBAA, Bilal BOUZIANE	Algeria, Algeria, Algeria	Submission 407
17	Valorization of Oil Sludge through Anaerobic	RIHANI Rachida, ALLOUNE Rhiad,	Algeria, Algeria,	Submission 408

	Co-Digestion with Cow Manure	CHEKNANE Benamar, CHAIBA Sidahmed, BENGAOUA Djamila, LOUCIF Sara	Algeria, Algeria, Algeria, Algeria	
18	Numerical development of thermal strain response of construction element by rheological model	Farid Bouziadi, Nefissa Belkacem, Bensaid Boulekbache, Mostefa Hamrat, Mohamed Ladjel	Algeria, Algeria, Algeria, Algeria, Algeria	Submission 413
19	Balancing Biodegradation and Stabilization through C:N Ratio Optimization in Date Palm Waste Composting	Abdelouahad Ouali, Abderrazak Mabrouki, Youssouf Ouzani, Houda Boutalbi, Fatima Hiouani, Djamila Madani, Chafia Tir	Algeria, Algeria, Algeria, Algeria, Algeria, Algeria, Algeria	Submission 414
20	Optimization of Nutrient-Enriched Edible Bowl Formulation Using Response Surface Methodology	Ankita Mondal, K. R. Jethva	India, India	Submission 415
21	Optical Rotary Encoder Accuracy Under Vibrations: Simulation–Experimental Parameter Correspondence	Violeta Krcheva, Stojance Nusev, Miša Tomić	North Macedonia, North Macedonia, Serbia	Submission 417
22	The Role of Islamic Finance for Startups and Self-Employed Enterprises: Banking and Tax Incentives	ZEROUKI Bouzned	Algeria	Submission 420
23	Material Identification of High-Speed Tool Steels via Reverse Engineering and Mechanical Testing	Mustafa Timur	Turkey	Submission 424
24	DFT Study of Ti3C2Tx/Co Multilayers for Electromagnetic Absorption and Infrared Stealth	R. Masrour	Morocco	Submission 425
25	DFT Study of Manganese-Based Oxide Cathodes for Aqueous Zinc-Ion Batteries	R. Masrour	Morocco	Submission 426
26	Spin-Dependent Optical Properties and Excitations in Layered CrPS4: A DFT Study	R. Masrour	Morocco	Submission 427
27	Experimental Investigation of Rate-Dependent Tensile Behavior and Axial Fatigue Life Performance of Injection-Molded PPGF30	Hamdi Kuleyin	Turkey	Submission 429
28	From Farm to Future: Pathways toward Sustainable Food System Transformation in Pakistan	Asif Ahmad, Zain Ul Iman , Asia Latif, Muhammad Saad Asif	Pakistan, Pakistan, Pakistan, Pakistan	Submission 437
29	Correlation Between Microstructural Refinement and Magnetic Properties of $Fe_{76}Nd_{16-x}Gd_xB_8$ Nanostructured Powders Prepared by Mechanical Alloying	Rachid Boughedaoui, Habib Achache, Mohamed Azzaz	Algeria, Algeria, Algeria	Submission 441
30	Deep Learning-Assisted Plasma Spraying for Hydrogen Storage Optimization	Boukhazani Fathi, Seddiki Mahdi	Algeria, Algeria	Submission 443
31	Numerical Optimization of Perovskite Solar Cell Performance Using SCAPS-1D	Boukhazani Fathi, Seddiki Mahdi	Algeria, Algeria	Submission 444
32	Detectability Analysis for Rectangular Linear Descriptor Systems via Normal System Transformation	Abed MAKRELOUFI	Algeria	Submission 447
33	Artificial Intelligence-Driven Prediction of Electromagnetic Performance of ZnO-Based Microstrip Patch Antennas	Amine bayoud, Abderrahim raidou	Morocco, Morocco	Submission 454
34	Connectionist vs. Kernel-Based Learning for Subsurface Residual Stress Mapping in Dry Turned Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 464
35	Non-Linear vs. Linear Machine Learning Architectures for Machining-Induced Subsurface Residual Stress Mapping in Inconel 718: Artificial Neural Networks vs. Linear SVM	Zaki Abdelfetah ABED	Algeria	Submission 465
36	Artificial Neural Networks vs. Rational Quadratic Gaussian Process Regression for Subsurface Residual Stress Profiling in Turned Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 466
37	Experimental and Numerical Analysis of Machining-Induced Residual Stresses in Hard Turning of Inconel 718 Using DEFORM-3D	Zaki Abdelfetah ABED	Algeria	Submission 467
38	Predictive Modeling of Residual Stresses in Hard Turned Inconel 718: A Comparative Study of 3D FEA Simulation and Multiple Linear Regression	Zaki Abdelfetah ABED	Algeria	Submission 468
39	Intelligent Prediction of Surface Integrity in Hard Turning of Inconel 718 Using ANN and DEFORM-3D	Zaki Abdelfetah ABED	Algeria	Submission 469
40	Linear SVM and FEM of Machining-Induced Residual Stresses in Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 470
41	Cubic Support Vector Machine Regression and FEM of Residual Stresses in Turning Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 471

42	Quadratic Support Vector Machine Regression and 3D FEA Modeling of Residual Stresses in Machined Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 472
43	Fine Gaussian Support Vector Machine Regression for High-Precision Residual Stress Prediction in Hard Turned Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 473
44	Medium Gaussian SVM for Balanced Residual Stress Mapping in Hard Turned Inconel 718: A Comparative Study with 3D FEA	Zaki Abdelfetah ABED	Algeria	Submission 474
45	A Comparative Study of ANFIS and 3D FEA Modeling for Predicting Surface Integrity in Hard Turned Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 475
46	Rational Quadratic GPR and 3D FEA for Non-Parametric Subsurface Stress Profiling in Hard Machined Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 476
47	Squared Exponential GPR and 3D FEA for Residual Stress Profiling in Turning Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 477
48	Surface Integrity Characterization of Turned Inconel 718 Using Matérn 5/2 Kernel GPR and DEFORM-3D Simulations	Zaki Abdelfetah ABED	Algeria	Submission 478
49	Non-Smooth Subsurface Stress Gradient Modeling in Hard Turned Inconel 718 via Exponential GPR	Zaki Abdelfetah ABED	Algeria	Submission 479
50	Hybrid Neuro-Fuzzy (ANFIS) vs. Artificial Neural Network (ANN) Modeling of Machining-Induced Residual Stresses in Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 480
51	ANFIS vs. Cubic Support Vector Machines: A Comparative Integrity Assessment of Turned Inconel 718 Surfaces	Zaki Abdelfetah ABED	Algeria	Submission 481
52	ANFIS vs. Linear Support Vector Machines: Performance Benchmarking for Residual Stress Prediction in Turned Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 482
53	Comparative Assessment of Hybrid Neuro-Fuzzy (ANFIS) and Multi-Scale Kernel (Rational Quadratic GPR) Models for Machined Inconel 718	Zaki Abdelfetah ABED	Algeria	Submission 483
54	A Review on the Remediation of Textile Dye-Containing Wastewater	Nazli Turkten, Yunus Karatas	Turkey, Turkey	Submission 484
55	Detection of performance degradation in aircraft engines with low pressure turbine vibration parameter by using artificial neural network and fuzzy logic	Bülent KURT, Mustagime Tülin YILDIRIM	Turkey, Turkey	Submission 496
56	Effect of Geocell Height on the Mechanical Behavior of Sand	Asseli Aymen Elouanas, Seif eddine Khadraoui	Algeria, Algeria	Submission 499
57	Quantitative Melissopalynological Profiling of 79 Honey Samples from the Ouarsenis Region, Northwestern Algeria	Fatima Hacene, Mounia Homrani, Abdelkader Homrani, Abdelkader Elamine Dahou	Algeria, Algeria, Algeria, Algeria	Submission 500
58	Autonomous Trajectory-Tracking Missions for Quadrotor UAVs: Implementation and Validation of Diverse Control Architectures	Marwa Gaidi, Soufyane Chekroune, Mokhtar Zerikat, Bouziane Meliani, Bouziane Ghoual	Algeria, Algeria, Algeria, Algeria, Algeria	Submission 501
59	Development of a strain-based rectangular membrane element for static and free vibration analysis	Sifeddine Abderrahmani	Algeria	Submission 505
60	HeliobiotracK (HBT): An Innovative Precision Livestock Farming Solution	BENAMARA Nasria	Algeria	Submission 509
61	Numerical Study of Flow and Heat Transfer in Conjugate Convection of a Micropolar Fluid	Sid Ahmed Alliche, Khaled Yahia Mahammed, Ridha Alliche	Algeria, Algeria, Algeria	Submission 511
62	Biocontrol Potential of Trichoderma afroharzianum Against Botrytis cinerea Causing Chocolate Spot of Faba Bean in Algeria	BOUANAKA Hamza, HARRAT Wahiba, SADRATI Nouari, BOUSSAHA Saoussene	Algeria, Algeria, Algeria, Algeria	Submission 512
63	Fabrication and Surface Characterization of TiC Coating on ST45 Steel via Electrodeposition	Tural BABAYEV, Halit SÜBÜTAY	Turkey, Turkey	Submission 517
64	Türk Enerji Birliği Perspektifinde Türk Devletleri Arasında HVDC Elektrik Enterkonneksiyonu: Teknik ve Stratejik Bir Değerlendirme	Turan ÇAKIL	Turkey	Submission 520
65	Deep Learning-Enhanced GIS for Geo-Tourism: Assessing the Educational and Economic Profitability of Gaberoun, Southern Libya	Asma Farhat Mohammed, Zaynab. M. Alfirgani, Llahm Omar Ben Dalla, Mansour Essgaer, Ebtisam Fakroun, Asma Agaal	Libya, Libya, Turkey, Libya, Libya, Libya	Mail Submission 6
66	Examination Monitoring System for	Ojo Olanrewaju, Taiwo Kehinde	Nigeria, Nigeria,	Submission 385

	Malpractice Detection and Evidence Logging	Ezekiel, Yekini Nureni, Olukumoro S. Olugbenga	Nigeria, Nigeria	
67	Understanding Consumer Behavior in the Wake of Cesarean Birth Trauma: A Pilot Study from Urban Pakistan	Shahid Mahmood, Raja Irfan Sabir	Pakistan, Pakistan	Submission 393
68	Expression of lncRNAs and circRNAs and Circadian Regulation of GH	Noor N. Abdullateef, Owayes M. Hamed	Iraq, Iraq	Submission 397
69	Regulatory Impact of Non-Coding RNAs on HK2 Gene Expression in Breast Cancer Patients: A Molecular Analysis	Safaa O. Yaarob, Owayes M. Hamed	Iraq, Iraq	Submission 398
70	Multiscale Characterization of Typha latifolia and Agave americana Fibres for Sustainable Composite Applications	Khalil Zouaoui, Mohamed-Saïd Chebbah, Salah Amroune	Algeria, Algeria, Algeria	Submission 411
71	Climate Change and the Spatial Dynamics of Poverty: A Geographical Study of Vulnerability in Agricultural Economy	OLANIYI Oladiran Paul, Awoyemi Olakunle Kolawole	Nigeria, Nigeria	Submission 439
72	Comparative Dynamic Behavior Analysis of Functionally Graded Beams with Power-Law and Sigmoidal Material Gradations: A Vibration Stability and Modal Characteristic Investigation	OUADAH Abbes, DRAI Ahmed	Algeria, Algeria	Submission 445
73	Critical Review of the Algerian Specifications for Private Hospital Establishments: Challenges of Adapting to New Architectural, Technological, and Healthcare Requirements	Alia Hassina	Algeria	Submission 452
74	SOME PROPERTIES OF THE CLASS (A, N) POWER-HYPONORMAL OPERATORS	Bekai Djilali, Djilali Laid, Benali Abdelkader, Hadj kaddour Tayeb	Algeria, Algeria, Algeria, Algeria	Submission 491
75	Leading AI-Enabled Process Reengineering: A Qualitative Study of Strategic Change and Leadership in the Intelligent Organization	Hartini Ahmad, Zulkifli Mohamed Udin, Jeniza Jamaluddin, Siti Hadijah Zulkifly	Malaysia, Malaysia, Malaysia, Malaysia	Submission 503
76	Stomatal Resistance and Ecophysiological Adaptation of Holm Oak (Quercus ilex L.) to Thermal Stress under Semi-Arid Mediterranean Conditions	HAICHOUR Rima, KHAZNADAR Mouna, Missaoui Khaled, BELDJAZIA Amina	Algeria, Algeria, Algeria, Algeria	Submission 514
77	Numerical Analysis of MHD Double-Diffusive Natural Convection in a CNT–Water Nanofluid-Filled Quadrantal Enclosure	Kamel Zitouni	Algeria	Submission 521
78-80	Inverse Gas Chromatography as a Tool for Surface and Thermodynamic Characterization	Ayse ERDOGAN CAKAR	Turkey	Submission 451
81	Altitude Estimation From Flight Data Recorder By Using Artificial Neural Network	Sancak Demiryürek, Mustagime Tülin YILDIRIM	Turkey, Turkey	Submission 519

Table List of Content

Content of Full Papers

Page No	Title	Authors	Country	Submission ID
1-4	MEKANİK SİSTEMLER İÇİN TERMAL AĞ MODELİ ÜRETEN MATLAB KODUNUN GELİŞTİRİLMESİ	Osama ELJOWHARY, Recep EKİCİLER	Turkey, Turkey	Mail Submission 1
5-9	The Effect of Silane-Modified Reinforced Epoxy on Co-Cured Single Lap Carbon-Fiber Composite Joints	İbrahim Erboğa, Mehmet Çağatay Akbolat, Ferhat Yetiş	Turkey, Turkey, Turkey	Mail Submission 2
10-15	Geological Constraints on the Hydrogen Economy in 2026: Storage Technology and Natural Hydrogen	Nazan YALÇIN ERİK	Turkey	Mail Submission 3
16-21	Artificial Intelligence in Petroleum Pollution Monitoring: Remote Sensing, Detection and Trajectory Modelling	Nazan YALÇIN ERİK	Turkey	Mail Submission 4
22-27	Detection of genes and some biochemical variables that related to chronic diseases	Zainab M. Fahmy, Zainab M. Rifat, Noor M. Atarbashi	Iraq, Iraq, Iraq	Mail Submission 5
28-35	Jeotermal Sularda Yüksek Br Varlığı: Akhüyük-Ereğli Jeotermal Alanı	Ayşen Davraz	Turkey	Submission 394
36-42	A Critical Evaluation of Total Electron Content (TEC) Anomalies in the Earthquake-Ionosphere Relationship in the Context of Geomagnetic and Solar Activity	Fuat Başçiftçi	Turkey	Submission 395
43-52	Unraveling the Molecular Basis of Gatekeeper Resistance: Comparative Molecular Docking and Dynamics Simulation of Erlotinib and Osimertinib against T790M-Mutant EGFR	Ömer Dilek, Tolga Acar Yeşil, Tahir Tilki	Turkey, Turkey, Turkey	Submission 403
53-56	The Importance of Medicinal and Aromatic Plants	Betül Gıdık	Turkey	Submission 404
57-64	A Comparative Performance Analysis of the YOLOv5s and YOLOv7-tiny Architectures in Multi-Flame Type Classification for Embedded Systems	Nazire Zeynep AYTEN, İlker GİDER, Sedat METLEK	Turkey, Turkey, Turkey	Submission 409
65-72	Retrieval-Augmented Generation (RAG) Sistemlerinin Karşılaştırmalı Analizi ve Uygulama Mimarileri	Hakan Yüksel	Turkey	Submission 412
73-78	DFT Optimization, Structural Validation, and In Silico Docking Study of a Coumarin Derivative on COX-2	Alpaslan Bayrakdar, İsmail Keleş	Turkey, Turkey	Submission 416
79-84	Evaluation of the Usage Fastness of Polyester/Viscose Blend Fabrics After Dyeing with Dyes of Different Reactivity	Aslıhan Koruyucu, Hatice Er	Turkey, Turkey	Submission 418
85-90	Epidotun İçeriğinin FTIR Spektroskopisi Kullanılarak Belirlenmesi	Ayşe Didem Kılıç, Ebubekir Sıddık İnanlı	Turkey, Turkey	Submission 419
91-97	GÖRME ENGELLİ BİREYLER İÇİN AKILLI BASTON TASARIMI	Ahmet TATAR, Adem KORKMAZ	Turkey, Turkey	Submission 421
98-103	Influence of Surcharge Loading on the Deformation and Shear Performance of Retaining Structures	Yesim Tuskan	Turkey	Submission 422
104-110	Consistency-Corrected Traveling Waves and Modulation Instability of the Shynaray-IIA Equation	Melike Kaplan	Turkey	Submission 423
111-116	Effects of Sodium Nitroprusside Applications on In Vitro Shoot Propagation and Development of Green Rose (Rosa chinensis var. 'Viridiflora')	Heydem EKİNCİ, Necla ŞAŞKIN, Bekir Erol AK	Turkey, Turkey, Turkey	Submission 428
117-122	Derin Öğrenme Kullanarak 3B Baskılı İletken PLA Çekirdek-Kabuk Kompozitlerinin Empedans Tabanlı Gerçek Zamanlı Kuvvet Tahmini	Dursun Ekmekci, Emrah Kaplan	Turkey, Germany	Submission 430
123-130	Samed Behrengi'nin Öykü ve Masallarında Yer Alan Eğitim Unsurlarının Çocuk Edebiyatı Bağlamında İncelenmesi	Kübra YILDIRIM CAN, Havvaana KARADENİZ	Turkey, Turkey	Submission 431
131-138	Green Building Optimization: Implementing AI-Driven Climate-Responsive Models to Lower Structural Energy Consumption	Murat Eyvaz	Turkey	Submission 432
139-147	Smart Distribution: Using IoT Sensors and Machine Learning for Real-Time Leak Detection and Loss Prevention	Murat Eyvaz	Turkey	Submission 436
148-154	GÜVENLİK HİZMETLERİNDE SÜRDÜRÜLEBİLİR İŞ GÜCÜ PLANLAMASI: HEDEF PROGRAMLAMA İLE BİR VARDİYA ÇİZELGELEME	Gülsüm Özge Coşkun, Günanç Karaca, Hacı Mehmet Alakaş	Turkey, Turkey, Turkey	Submission 438

	UYGULAMASI			
155-161	Digital Acquisition Requirements for Accurate Damping Identification from Free-Vibration Responses	Yusuf HAMİDA EL NASER	Turkey	Submission 440
162-164	Climate Change and Increasing Electricity Demand in Eastern Turkey: Current Situation, Trends and Future Predictions	İshak Parlar, Onur Silahtar	Turkey, Turkey	Submission 446
165-172	Evaluation of the Effectiveness of Random Forest and Other Classification Algorithms in Predicting Heart Disease Using SHAP Analysis	Halit Çetiner	Turkey	Submission 448
173-180	Design and Simulation of an Intuitionistic Takagi–Sugeno Fuzzy Controller for DC Motor Speed Control	Onur SİLAHTAR, İshak PARLAR	Turkey, Turkey	Submission 449
181-188	In Vitro Micropropagation of Mango (Mangifera indica L.) and Factors Affecting Culture Success	Fatma ALAN, Müttalip GÜNDOĞDU, Ayşen Melda ÇOLAK	Turkey, Turkey, Turkey	Submission 457
189-193	New Rough Approximations Induced by Families Generated via Local Closure	Ferit Yalaz	Turkey	Submission 458
194-200	Artificial intelligence in the function of improving the educational process	Nebojša Denić, Ivana Bulut Bogdanović, Saša Mihajlović, Momir Milić, Stefan Milić	Serbia, Serbia, Serbia, Serbia, Serbia	Submission 460
201-208	Vessel Telemetry Data Analytics: DBSCAN-Driven Route Clustering and Isolation Forest-Based Anomaly Detection	Sertaç YAMAN, Mehmet Latif LEVENT, Halit HÜLAKO	Turkey, Turkey, Turkey	Submission 461
209-217	MSFOA-Based Hyperparameter Optimization for Machine Learning Models in Dry Bean Variety Classification	Harun Akbulut, Muhammet Emin Sahin, Orhan Er	Turkey, Turkey, Turkey	Submission 462
218-230	Background-Induced Shortcut Learning in Cucumber Quality Classification: A Comparative Evaluation of ConvNeXt-Tiny and EfficientNetV2-S Architectures	Esra Yuce, Hasan Ulutas, H. İbrahim Cosar, M. Emin Sahin	Turkey, Turkey, Turkey, Turkey	Submission 463
231-233	Some Applications of the Euler Equation in Control Theory	Muzaffer ATEŞ, Muhammet ATEŞ	Turkey, Turkey	Submission 485
234-236	Dinamik Programlama Yaklaşımı ile Kısa Mesafe Probleminin Çözümü	Muzaffer ATEŞ, Hakan ALAN	Turkey, Turkey	Submission 487
237-247	When Retrieval Helps — and When It Hurts: A Reliability Framework for Retrieval-Augmented Generation and Forecasting	Ali YILDIZ, Ali AKDAĞLI	Turkey, Turkey	Submission 488
248-254	ABD'de Haftalık COVID-19 Ölüm Sayılarının Tahmini: Makine Öğrenmesi ve Derin Öğrenme Modellerinin Deneysel Analizi	Anıl Utku	Turkey	Submission 489
255-262	Sequence Homology Inflates Reported Accuracy in Antimicrobial Peptide Prediction: A Machine Learning and Deep Learning Benchmark	Erkan Caner Ozkat, Gozde Yalcin Ozkat	Turkey, Turkey	Submission 490
263-271	Next-Generation Marine Fender Systems: A Critical Review and Conceptual Feasibility Assessment of Additively Manufactured TPU Lattice Structures	Tolga Berkay Şirin, Ahmet Can Mutlu, Utku Cem Karabulut, Ramazan Özkan Yıldız, Halim İbrahim Erbaş, Oğuzhan Der	Turkey, Turkey, Turkey, Turkey, Turkey, Turkey	Submission 492
272-278	Influence of Build and Lattice Orientation on the Performance of Additively Manufactured TPU Boat Fenders: A Case Study on the Transition Towards a Sustainable Maritime Industry	Tolga Berkay Şirin, Utku Cem Karabulut, Ahmet Can Mutlu, Ramazan Özkan Yıldız, Halim İbrahim Erbaş, Oğuzhan Der	Turkey, Turkey, Turkey, Turkey, Turkey, Turkey	Submission 493
279-293	Effects of Introducing a Steel Sheet Between Two Rows of Brick Masonry Wall on Modal Parameters	Furkan Günday	Turkey	Submission 494
294-303	ESG Integration in Pistachio Production, Processing and Packaging: A Climate-Resilient Farm-to-Market Framework	GEORGIOS J. ZAKYNTHINOS, Bekir Erol AK	Greece, Turkey	Submission 495
304-308	Dietary Exposure and Health-Risk Reassessment of Toxic Metals in Black Tea for a High-Consumption Population: A Türkiye-Focused Analysis of an Open-Access Dataset	Fatih Özkalaycı, Rüveyda Esra Özkalaycı	Turkey, Turkey	Submission 497
309-316	Proje Bazlı Üretimde Ortak Malzemeler İçin Önceliklendirme Modeli	Ezgi GÜL, M. Hanefi CALP	Turkey, Turkey	Submission 504
317-324	Horse Chestnut in the Urban Landscape: Ecological Value and the Growing Threat of Cameraria ohridella	Seray ÖZDEN KELEŞ	Turkey	Submission 506
325-328	Some Oil Crops and Their Importance	Betül Gıdık	Turkey	Submission 507
329-338	Comparison of Quality Characteristics of Natural, Hibiscus Extract and Artificial Lemonade Concentrates: Functional and Physicochemical	Hacer GÜRKAN, Nida Nur DÜZENLİ, Muayıd ALMUHAMED	Turkey, Turkey, Turkey	Submission 508

339-346	Hydrogen Utilization in Compression Ignition Diesel Engines: Opportunities, Challenges and Future Perspectives	Yakup Daşdemirli, Nebi Yelegen	Turkey, Turkey	Submission 516
347-354	Waste Cooking Oil Biodiesel in Diesel Engines: A Critical Review on Manufacturing Process, Engine Efficiency, Emissions, Material Compatibility, and Circular Economy Opportunities	Yakup Daşdemirli, Nebi Yelegen	Turkey, Turkey	Submission 518

COMMITTEE

ORGANIZING COMMITTEES

Assoc. Prof. Admir Jance, European University of Tirana
Assoc. Prof. Elif Küçük Durur, Atatürk University
Assoc. Prof. Nermin Kişi, Zonguldak Bülent Ecevit University
Assoc. Prof. Ülhak Çimen, Atatürk University
Dr. Krisztina Czakoova, J. Selye University
Dr. Ondrej Takac, J. Selye University
Dr. Ladislav Vegh, J. Selye University

SCIENTIFIC COMMITTES

Prof. Dr. Abdurazzag Ali Aburas, University of Kwazulu-Natal
Prof. Dr. Adnan Sözen, Gazi University
Prof. Dr. Amit Soni, Manıpal University
Prof. Dr. Anis Koubaa, Prince Sultan University
Prof. Dr. Carlos A. Coelho, Universidade Nova de Lisboa
Prof. Dr. Josep Guerrero, Aalborg University
Prof. Dr. Juan Vasquez, Aalborg University
Prof. Dr. Laszlo T. Koczy, Budapest University of Technology and Economics
Prof. Dr. Mohammad Ashraf Gondal, King Fahd University
Prof. Dr. Murat Yakar, Mersin University, Turkey
Prof. Dr. Muhammad Rashid, University of West Florida
Prof. Dr. Saad Mekhilef, Malaya University
Prof. Dr. Salim Ibrir, King Fahd University
Prof. Dr. Subarna Shakya, Tribhuvan University
Prof. Dr. Suzuki Hisashi, Chuo University
Prof. Dr. Usha Mehta, Institute of Technology, Nirma University
Prof. Dr. Vagif Gasimov, Azerbaijan Technical University
Prof. Dr. Weidong Zhu, University of Maryland
Assoc. Prof. Ali Hakan Işık, Burdur Mehmet Akif Ersoy University
Assoc. Prof. Amita Nandal, Manıpal University
Assoc. Prof. Aslan İnan, Yıldız Teknik University
Assoc. Prof. Atif Iqbal, Qatar University
Assoc. Prof. Elif Küçük Durur, Atatürk University
Assoc. Prof. Francesco Cottone, University of Perugia
Assoc. Prof. Farid Melgani, University of Trento
Assoc. Prof. Halit Öztekin, Sakarya Applied Science University
Assoc. Prof. Hamdy Ziedan, Assiut University
Assoc. Prof. Harun Kemal Öztürk, Pamukkale University
Assoc. Prof. Jiri Pela, Brno University of Technology
Assoc. Prof. Merih Palandöken, İzmir Katip Çelebi University
Assoc. Prof. Nermin Kişi, Zonguldak Bülent Ecevit University
Assoc. Prof. Rita Ismailova, Kyrgyz Turkish Manas University
Assoc. Prof. Sadia Samar Ali, King Abdul-Aziz University
Assoc. Prof. Ülhak Çimen, Atatürk University
Assoc. Prof. Tanmay De, National Institute of Technology Durgapur
Assoc. Prof. Zeashan Khan, Air University
Dr. Ali Almisreb, International University of Sarajevo
Dr. Amina Othmani, University of Monastir-Tunisia
Dr. Arti Jain, Jaypee Institute of Information Technology
Dr. Arun Kumar Verma, Malaviya National Institute of Technology Jaipur
Dr. Aziz Shah, University of Glaskow
Dr. Bal Chand Nagar, ECE National Institute of Technology Patna
Dr. Baoze Wei, Aalborg University
Dr. Bohumil Brtník, University of Pardubice
Dr. Diptiranjan Behera, The University of the West Indies at Mona
Dr. Fayçal BENSAALI, Qatar University
Dr. Fugui Qi, Fourth Military Medical University
Dr. Magdi S. Mahmoud, King Fahd University
Dr. Mariya Ouaissa, Moulay Ismail University
Dr. Merdaci Slimane, University of -Sidi Bel Abbés, Algeria
Dr. Moharram Challenger, Antwerp University
Dr. Muhammad Sultan, Zakariya University
Dr. Lesya Anishchenko, Bauman Moscow State Technical University
Dr. R.A. Ilyas, Universiti Teknologi Malaysia
Dr. Sadina Gagula Palalić, International University of Sarajevo
Dr. Stefania Tomasiello, University of Tartu
Dr. Syed Omer Gilani, National University of Sciences and Technology (NUST)
Dr. Tarık Namas, International University of Sarajevo
Dr. Uduak A. Umoh, University of Uyo
Dr. Yajuan Guan, Aalborg University

5TH INTERNATIONAL CONFERENCE ON RECENT AND INNOVATIVE RESULTS IN ENGINEERING AND TECHNOLOGY ICRIRET 2026

AUGUST 25-26, 2026

05.09.2026

25-26 Ağustos 2026 tarihlerinde MEET üzerinden çevrimiçi olarak gerçekleştirilen 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026 konferansı akademik teşvik yönetmeliğinin 9. Maddesine istinaden "Tebliğlerin sunulduğu yurt içinde veya yurt dışındaki etkinliğin uluslararası olarak nitelendirilebilmesi için Türkiye dışında en az beş farklı ülkeden sözlü tebliğ sunan konuşmacının katılım sağlaması ve tebliğlerin yarıdan fazlasının Türkiye dışından katılımcılar tarafından sunulması esastır." kriterlerini sağlamaktadır. Toplam 127 adet bildirinin yer aldığı kongre iki gün boyunca çevrimiçi olarak gerçekleştirilmiştir.

Türkiye dışından toplam 15 farklı ülkeden (Fas, Cezayir, Pakistan, Kıbrıs, Yunanistan, Almanya, Birleşik Krallık, Hindistan, Nijerya, Bangladeş, Irak, Kuzey Makedonya, Sırbistan, Malezya, Libya) katılım sağlanmış olup, 127 adet bildirinin 70 (%55,11) tanesi yabancı katılımcı tarafından sunulmuştur.

Kongremize ilginiz için teşekkür ederiz.

Saygılarımızla,



All Sciences Academy

05.09.2026

25-26 Ağustos 2026 tarihlerinde MEET üzerinden çevrimiçi olarak gerçekleştirilen 5th International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2026 Konferansı, 2025 MART DÖNEMİ DOÇENTLİK BAŞVURU ŞARTLARI yönetmeliğinin 2. bölümde yer alan Uluslararası Bilimsel Toplantı tanımlaması "Farklı ülkelerden bilim insanlarının bilim kurulunda bulunduğu ve sunumların bilimsel ön incelemeden geçirilerek kabul edildiği toplantı." şeklindedir.

5TH ICRIRET 2026 konferansı bilimsel komitesinin çoğunluğu yabancı bilim insanlarından oluşmaktadır. Aşağıdaki link ile bilimsel komite incelenebilir:

https://as-proceeding.com/index.php/icriret/committe

Ayrıca konferansımızda hakem değerlendirmesi olup hakem değerlendirmesi olumlu sonuçlanmadan ücret talebinde bulunulmamaktadır.

2024 MART DÖNEMİ DOÇENTLİK BAŞVURU ŞARTLARI yönetmeliğinin 8. bölümde yer alan Bilimsel Toplantı tanımlamasının b bendindeki "Diğer Uluslararası/Ulusal bilimsel toplantının düzenlenme komitesinde, kurum/tüzel kişilik/karar organı tarafından resmi olarak görevlendirilmiş üniversite/enstitü/bilimsel kurum/branş derneği akademisyen temsilcisi bulunması zorunludur." şeklindedir.

5TH ICRIRET 2026 konferansı düzenleme komitesinde farklı kurumlar tarafından görevlendirilmiş birden fazla üniversiteden akademik temsilciler bulunmaktadır. Aşağıdaki link ile düzenleme komitesindeki akademik temsilcilerin görevlendirme yazıları incelenebilir:

https://drive.google.com/file/d/1eZalxRl8q6-eXATZzwUraH9KOhtQzFho/view

Ayrıca görevlendirme yazıları ilgili konferansın özet ve tam metin kitapçıklarında yer almaktadır.

Kongremize ilginiz için teşekkür ederiz.

Saygılarımızla,



All Sciences Academy

04.09.2023

LETTER OF ASSIGNMENT

We confirm that Ing. Ondrej Takáč, PhD. working at J. Selye University takes part in the organization committee of the conferences organized by of All Sciences Academy.

Confirmation of the Vice Dean of the Faculty of Economics and Informatics

Name of Responsible Person and Signature:

Name and Surname Sign:

PhDr. Imrich Antalík, PhD. _____

Date: 04.09.2023

UNIVERZITA J. SELYEHO
SELYE JÁNOS EGYETEM
FAKULTA EKONÓMIE A INFORMATIKY
DEKANÁT
Bratislavská 167/21, 945 01 Komárno

04.09.2023

LETTER OF ASSIGNMENT

We confirm that PaedDr. Ladislav Végh, PhD. working at J. Selye University takes part in the organization committee of the conferences organized by of All Sciences Academy.

Confirmation of the Vice Dean of the Faculty of Economics and Informatics

Name of Responsible Person and Signature:

Name and Surname Sign:

PhDr. Imrich Antalík, PhD. _____

Date: 04.09.2023

Universiteti Europian i Tiranës

16.08.2023

LETTER OF ASSIGNMENT

We confirm that Prof. Assoc. Dr. Admir Jançe, who is employed full-time as a lecturer, as well as holding the position of Head of the Imaging Department and Vice Dean of the Faculty of Technical Medical Sciences, takes part in the organization committee of the conferences organized by of All Sciences Academy.

Confirmation of the Rector/Dean/Vice Dean/Head of Department

Name of Responsible Person and Signature:



Rector: Prof. Dr. Tonin GJURAJ

_ _

Date: 16.08.2023

15.09.2023

İlgili Makama

Atatürk Üniversitesi İletişim Fakültesinde görev yapan Doçent Doktor Ülhak Çimen'in All Sciences Academy yayınevi ile beraber düzenlenmesi planlanan konferanslarda ve organizasyonlarda düzenleme veya bilim kurullarında yer alması uygun görülmüştür.

Gereğini rica ederim.

Ad ve Soyad Prof. Dr. Adem YILMAZ Dekan Tarih 15.09.2023	İmza: Prof.Dr. Adem YILMAZ DEKAN

15.09.2023

İlgili Makama

Atatürk Üniversitesi İletişim Fakültesinde görev yapan Doçent Doktor Elif Küçük DURUR'un All Sciences Academy yayınevi ile beraber düzenlenmesi planlanan konferanslarda ve organizasyonlarda düzenleme veya bilim kurullarında yer alması uygun görülmüştür.

Gereğini rica ederim.

Ad ve Soyad Prof. Dr. Adem YILMAZ Dekan Tarih 15.09.2023	İmza: Prof.Dr. Adem YILMAZ DEKAN

04.09.2023

LETTER OF ASSIGNMENT

We confirm that PaedDr. Krisztina Czakóová, PhD. working at J. Selye University takes part in the organization committee of the conferences organized by of All Sciences Academy.

Confirmation of the Vice Dean of the Faculty of Economics and Informatics

Name of Responsible Person and Signature:

Name and Surname Sign: _____

PhDr. Imrich Antalík, PhD.

Date: 04.09.2023

T.C.
ZONGULDAK BÜLENT ECEVİT ÜNİVERSİTESİ REKTÖRLÜĞÜ
Çaycuma Meslek Yüksekokulu Müdürlüğü

Sayı : E-84604848-903.07.04-371012 01.11.2023
Konu : Bilimsel Toplantı Görevlendirme

İLGİLİ MAKAMA

İlgi : Öğretim Üyesi Doç. Dr. Nermin KİŞİ'nin 30.10.2023 tarihli dilekçesi.

Yüksekokulumuz Yönetim ve Organizasyon Bölümü Öğretim Üyesi Doç. Dr. Nermin KİŞİ'nin "All Sciences Academy Yayınevi"nin düzenleyeceği bilimsel toplantıların düzenleme ve bilim kurullarında görev almasında Yüksekokulumuzca bir sakınca bulunmamaktadır. İş bu belge ilgilinin isteği üzerine tanzim edilmiştir.

Bilgilerinizi ve gereğini rica/arz ederim.

Doç. Dr. Hikmet YAZICI
Meslek Yüksekokulu Müdürü

Bu belge, güvenli elektronik imza ile imzalanmıştır.

Adres : BEÜ Çaycuma Meslek Yüksekokulu, 67900, Çaycuma, Zonguldak
Telefon : (0372) 643 66 01 Faks:(0372) 643 66 04
e-Posta : caycumamyo@beun.edu.tr Web : https://cmyo.beun.edu.tr/
Kep Adresi : beun@hs03.kep.tr
Bilgi için : Tülay KEÇECİ
Unvanı : İşçi

5th International Conference on Recent and Innovative Results in Engineering and Technology

August 25-26, 2026: Konya, Turkey

© 2026 Published by All Sciences Academy

https://as-proceeding.com/index.php/icriret/home

Sequence Homology Inflates Reported Accuracy in Antimicrobial Peptide Prediction: A Machine Learning and Deep Learning Benchmark

Erkan Caner Ozkat*[1] and Gozde Yalcin Ozkat[2]

[1]Department of Mechanical Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdogan University, Rize, Turkiye (ORCID: https://orcid.org/0000-0003-0530-5439)

[2]Department of Bioengineering, Faculty of Engineering and Architecture, Recep Tayyip Erdogan University, Rize, Turkiye (ORCID: https://orcid.org/0000-0002-9689-2239)

*(erkancaner.ozkat@erdogan.edu.tr) Email of the corresponding author

Abstract – Machine learning is now routine for screening antimicrobial peptide candidates, and reported accuracies are high. This study asks how much of that accuracy depends on sequence homology between training and test data. A benchmark of 6288 mature peptides was assembled from curated UniProtKB annotation, using one extraction rule for both classes, so negative-set bias is excluded by construction. All 19,766,328 peptide pairs were aligned with local Smith-Waterman scoring. Partitions were drawn at 0.8 identity, the only threshold that clears a shuffled-sequence noise floor without chaining the set into one component. Twenty pipelines were then trained under two partitioning regimes with five seeds each. They comprise sixteen classical pipelines, three neural architectures and a control that only looks up the most similar training peptide. The peptide set proved strongly redundant, and a random partition left 62 per cent of test peptides with a close relative in training. Every trained pipeline lost accuracy when whole identity components were held out, by 0.039 in area under the curve on average, and the leading pipeline changed. The lookup control reached 0.958 under the random partition against 0.819 under the homology-aware one. Accuracy fell steadily as identity to the nearest training peptide decreased. Holding out whole greedy clusters, which is common practice, still left 87 per cent of held-out peptides with a close relative in training. Sequence-based peptide classifiers should therefore be partitioned by identity component, and test-to-training similarity should be reported alongside the headline metric.

Keywords – Antimicrobial Peptides, Data Leakage, Sequence Homology, Deep Learning, Benchmark Evaluation

I. INTRODUCTION

Bacterial antimicrobial resistance was associated with an estimated 4.95 million deaths in 2019 [1]. Antimicrobial peptides are short and mostly cationic sequences that act on microbial membranes [2]. They remain an attractive route to agents against resistant organisms, because they kill by physical disruption rather than through a single protein target [3]. Resistance therefore develops slowly against many of them. Laboratory screening of candidates is slow and costly, so computational triage has become a standard first step [4].

Sequence-based classifiers dominate that triage step. Early methods encoded a peptide as amino-acid or dipeptide composition, or as a block of biophysical descriptors. The encoded peptide was then passed to a random forest or a support vector machine [5]. Deep models followed, learning residue embeddings with convolutional and recurrent layers [6, 7]. The same architectures now drive genome mining and generative peptide design [8, 9]. Reported performance is uniformly high, and areas under the receiver operating characteristic curve above 0.95 are common.

Independent assessments do not support that reading. Web tools evaluated on unseen data lose a large part of their advertised accuracy [10]. A systematic comparison of published predictors found that their ranking depends on how the evaluation set was assembled [11]. One cause is already documented. The negative class in these benchmarks usually differs from the positive class in length and in composition,

which makes the task easier than it should be [12]. A second cause is generic to biological sequence data. Natural peptides fall into families whose members share ancestry [13]. A partition drawn at random therefore places close relatives on both sides of it. The same failure is well described for paired biological inputs [14] and for machine learning in genomics [15]. It is now treated as a leading source of irreproducible results [16].

This paper measures how much reported accuracy survives when the second cause is removed. A peptide set is assembled from primary annotation by a rule that is identical for both classes. That isolates the homology question from the negative-selection question. Every pair of peptides is then aligned, and the redundancy of the set is quantified against a composition-matched null. Two partitions are drawn. The first is the usual random partition. The second removes whole identity clusters, so no test peptide has a close relative in training. Twenty prediction pipelines are trained under both partitions.

The study contributes five things. It provides a reproducible peptide benchmark built from a single primary source by a symmetric rule. It calibrates the identity threshold against shuffled sequences rather than adopting a convention. It shows that a partition drawn on greedy clusters still leaves close relatives on both sides, and quantifies that gap. It reports a like-for-like comparison of classical and deep pipelines under both regimes. It relates accuracy to the identity between each test peptide and its nearest training neighbour. The last step turns a binary argument about splitting into a measured curve.

II. Materials and Method

A. Peptide dataset

Peptides were taken from the reviewed section of UniProtKB, which carries manually curated sequence features [17]. Two pools were downloaded. The positive pool held entries annotated with the Antimicrobial keyword. The negative pool held entries that carry at least one annotated mature peptide, but none of the antimicrobial, antibiotic, antiviral, fungicide, toxin or bacteriolytic keywords. Membrane-lytic venom peptides are often antimicrobial without being annotated as such. The toxin keyword is excluded from the negative pool for that reason.

One row was extracted per annotated mature peptide feature, using the same rule in both pools. Features with uncertain boundaries were skipped. So were peptides outside five to one hundred residues, and peptides containing a non-standard amino-acid symbol. Sequences seen in more than one entry were collapsed to a single row. Thirty sequences that appeared under both labels were discarded as ambiguous. The class label is therefore a property of the parent entry, and the two classes are matched on annotation type. Table 1 describes the resulting set of 6288 peptides.

Table 1. Composition of the peptide benchmark.

Class	Peptides	Median length	Length IQR	Source organisms	UniProt entries
Antimicrobial	1846	25	18-33	432	1587
Non-antimicrobial	4442	20	11-31	842	3011

B. Sequence identity and clustering

All 19,766,328 pairs of peptides were aligned with the local Smith-Waterman algorithm [18]. Alignments were scored with BLOSUM62 [19], using a gap penalty of eleven for opening and one for extension. Identity is the number of identical aligned residues, divided by the length of the shorter peptide. A short local hit inside a long peptide therefore scores low. Dipeptide word filters lose almost all pruning power at the thresholds relevant to short peptides. The matrix was computed exhaustively for that reason, rather than approximated.

Two grouping rules were then applied to that matrix. The first is the greedy longest-first rule used by CD-HIT, which reports how much of the set is redundant [13]. The second is the transitive closure, in which every pair at or above the threshold is placed in one component. Only the closure supports a partition. Greedy clustering compares each sequence with cluster representatives alone, so two sequences in different greedy clusters can still be near-identical.

The threshold was calibrated rather than assumed. Each sequence was shuffled, which preserves its length and composition while destroying its ancestry. Pairs of shuffled sequences were then aligned in the

same way. Identity below 0.6 is not distinguishable from chance for peptides of this length. The shuffled null links 1.7 per cent of pairs at 0.4 and 0.08 per cent at 0.6. That places a floor under any usable threshold.

The closure then places a second constraint from above. At 0.6 identity the closure chains 96.6 per cent of the peptides into a single component, so no partition at that threshold is possible. At 0.7 the largest component still holds 49.2 per cent. At 0.8 it holds 1.8 per cent, and the set breaks into 2782 components. A partition at 0.6 must therefore hold out either 3.4 per cent of the peptides or 96.6 per cent of them. Neither is a usable test set. The homology-aware partition operates at 0.8, which is the least strict threshold that both clears the noise floor and admits a partition.

C. Partitioning regimes

Two regimes were compared, each drawn five times with different seeds. The random regime draws a stratified twenty per cent of peptides as the test set. That is the usual practice in this literature. The homology-aware regime instead draws whole identity components until the same share is reached. Components are chosen so that the class balance of the held-out side tracks the full set. No test peptide can then share 0.8 identity with any training peptide, because such a pair would lie in one component. An inner validation set of twenty per cent of the remainder is drawn under the same rule as the outer split. Early stopping and threshold selection therefore never see the leakage that the outer split excludes.

D. Representations and models

Sixteen classical pipelines cross four encodings with four learners. The encodings are amino-acid composition, dipeptide composition, a block of twenty-five biophysical descriptors, and the concatenation of the first and third. The descriptor block covers length, mass, net charge at pH 7.4 and isoelectric point. It also covers mean hydropathy [20] and the largest sliding-window hydrophobic moment at a helical angle of one hundred degrees [21]. The aliphatic index, the instability index and the Boman index [22] complete the physical block, together with predicted secondary-structure fractions and ten residue-group fractions. The learners are logistic regression and a support vector machine with a radial basis kernel. A random forest [23] and gradient boosted trees [24] complete the set, each with balanced class weights.

Three neural architectures read the residue sequence directly through a learned embedding. The first stacks two convolutions and pools over positions. The second is a bidirectional long short-term memory network [25]. The third combines a convolutional front end with a recurrent body and attention pooling. Batches are sorted by length and padded to their own longest member. Three quarters of a fixed-width tensor would otherwise be padding.

A twentieth pipeline is included as a control, and it fits nothing. For each test peptide it finds the most similar antimicrobial peptide in training. It also finds the most similar non-antimicrobial one. The peptide is scored by the difference between those two identities. Any accuracy the control reaches is attributable to sequence similarity alone. It therefore measures how far a benchmark can be solved by lookup.

E. Training and evaluation

Every pipeline saw the same training peptides and was scored on the same test peptides. Neural models were trained for at most thirty epochs with the Adam optimiser. The learning rate was 0.001, the batch size was 64 and the loss was class-weighted. Training stopped early after five epochs without improvement in validation area under the curve. Decision thresholds were selected on the validation peptides by maximising the Matthews correlation coefficient, which is the recommended summary for imbalanced problems [26]. Test results are reported by area under the curve, average precision, the Matthews coefficient, sensitivity and specificity.

The pipeline was implemented in Python 3.12 with Biopython [27], scikit-learn [28], XGBoost and PyTorch [29]. The complete grid of two regimes, five seeds and twenty pipelines gives 200 fitted models. It runs in about 8 minutes on eight cores. Aligning all peptide pairs takes about eleven minutes on the same machine.

III. Results

A. The peptide set is strongly redundant

The first result concerns the data rather than any model. At the calibrated threshold of 0.6, the 6288 peptides collapse into 1948 identity clusters. So 69 per cent of the set is redundant. Only 14 per cent of peptides are singletons, and the largest family holds 107 members. Redundancy does not disappear at stricter thresholds. Even at 0.9 identity, 47 per cent of peptides still share a cluster with another peptide. Figure 1 shows the length distribution, the distribution of pairwise identity, and the way both redundancy and linkage track the threshold.

That redundancy passes straight into a random partition. Averaged over five random splits, 62 per cent of test peptides have a training peptide at 0.8 identity or above. That figure rises to 95 per cent at 0.6 identity. The mean identity to the nearest training peptide is 0.83. Under the homology-aware splits no test peptide reaches 0.8 identity with any training peptide, and the mean nearest identity falls to 0.66. The two regimes therefore pose different questions. The first asks whether a model recognises relatives of peptides it has seen. The second asks whether it recognises peptides unlike anything it has seen.

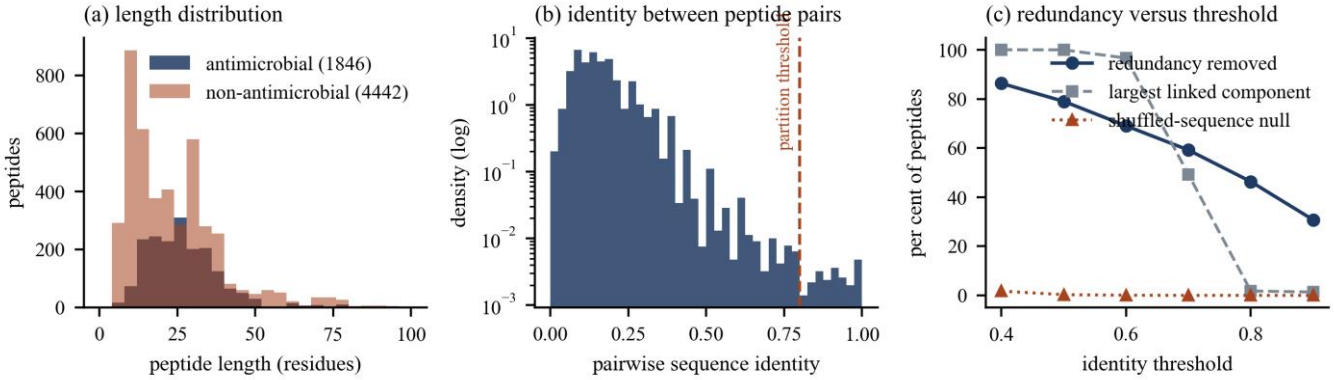


Fig. 1 Composition and redundancy of the peptide benchmark. Panel (c) shows redundancy, the share of peptides in the largest linked component, and the rate at which shuffled sequences of the same composition are linked.

B. Cluster-disjoint splitting is not homology-free

A partition drawn on greedy clusters does not deliver what it promises, and this is worth separating from the main result. Greedy clusters at 0.6 identity were held out whole, so no cluster appears on both sides. Even so, 87 per cent of held-out peptides still had a training peptide at 0.6 identity or above. The mean identity to the nearest training peptide was 0.67. Holding out greedy clusters at 0.8 leaves 23 per cent of test peptides with a relative at that level. Components remove both figures to zero by construction, which is why the benchmark uses them.

The reason is structural rather than incidental. Greedy assignment compares each sequence with cluster representatives alone. Two peptides that match no common representative can still be near-identical to each other, and they are then placed in different clusters. Any study that splits on clusters from a greedy tool inherits this gap. Reporting the achieved identity between test and training sequences would expose it immediately.

C. Accuracy under the two partitioning regimes

Table 2 reports every pipeline under both regimes. Figure 2 shows the same values as paired points. Under the random regime the best pipeline is DPC/RF, at an area under the curve of 0.970. Under the homology-aware regime the leader is HYBRID/RF, at 0.936. Averaged over the nineteen trained pipelines, the change of regime costs 0.039. The loss is present for every pipeline, and ranges from 0.008 to 0.138. The pooled paired test over all pipelines and seeds gives $p < 0.0001$.

The lookup control makes the mechanism explicit. That pipeline fits nothing, and only compares a test peptide with its most similar training neighbours. It still reaches 0.958 under the random regime. The best trained pipeline on the same splits reaches 0.970. Under the homology-aware regime the same control

falls to 0.819. Sequence similarity alone therefore accounts for most of what a random split measures. It accounts for much less of what a homology-aware split measures.

Model choice matters less than the partition. Within the sixteen classical pipelines, the regime explains 24.3 per cent of the variance in the area under the curve. The encoding explains 1.8 per cent and the learner 54.9 per cent. The seed contributes 6.5 per cent. The best deep architecture reaches 0.966 under the random regime and 0.923 under the homology-aware one. The best classical pipeline reaches 0.970 and 0.936. Deep models therefore fall behind their classical counterparts once relatives are removed.

Table 2. Test AUROC under both regimes, as the mean and standard deviation of five seeds. Delta is the accuracy lost when relatives are removed from training.

Pipeline	AUROC random	AUROC homology-aware	Delta	MCC random	MCC homology
NN-identity	0.958 ± 0.003	0.819 ± 0.030	0.138	0.783	0.506
HYBRID/RF	0.968 ± 0.007	0.936 ± 0.013	0.033	0.817	0.746
HYBRID/XGB	0.968 ± 0.009	0.933 ± 0.015	0.035	0.836	0.731
HYBRID/SVM	0.956 ± 0.011	0.929 ± 0.014	0.027	0.806	0.742
PHYSCHEM/RF	0.962 ± 0.008	0.926 ± 0.017	0.037	0.814	0.732
AAC/RF	0.966 ± 0.008	0.923 ± 0.010	0.042	0.796	0.688
PHYSCHEM/XGB	0.963 ± 0.009	0.923 ± 0.017	0.040	0.825	0.706
PHYSCHEM/SVM	0.947 ± 0.015	0.919 ± 0.018	0.028	0.792	0.715
AAC/XGB	0.963 ± 0.007	0.917 ± 0.019	0.045	0.793	0.681
AAC/SVM	0.948 ± 0.010	0.915 ± 0.016	0.033	0.760	0.677
DPC/RF	0.970 ± 0.004	0.911 ± 0.018	0.059	0.828	0.705
DPC/SVM	0.961 ± 0.008	0.904 ± 0.020	0.057	0.811	0.683
DPC/XGB	0.960 ± 0.003	0.901 ± 0.020	0.059	0.802	0.683
HYBRID/LR	0.889 ± 0.016	0.881 ± 0.023	0.008	0.634	0.635
PHYSCHEM/LR	0.884 ± 0.016	0.876 ± 0.023	0.009	0.632	0.647
AAC/LR	0.866 ± 0.014	0.856 ± 0.012	0.010	0.593	0.588
DPC/LR	0.916 ± 0.005	0.851 ± 0.021	0.065	0.680	0.549
CNN-BiLSTM	0.962 ± 0.008	0.923 ± 0.013	0.039	0.811	0.711
BiLSTM	0.956 ± 0.013	0.912 ± 0.022	0.044	0.795	0.697
CNN	0.966 ± 0.008	0.903 ± 0.016	0.063	0.818	0.675

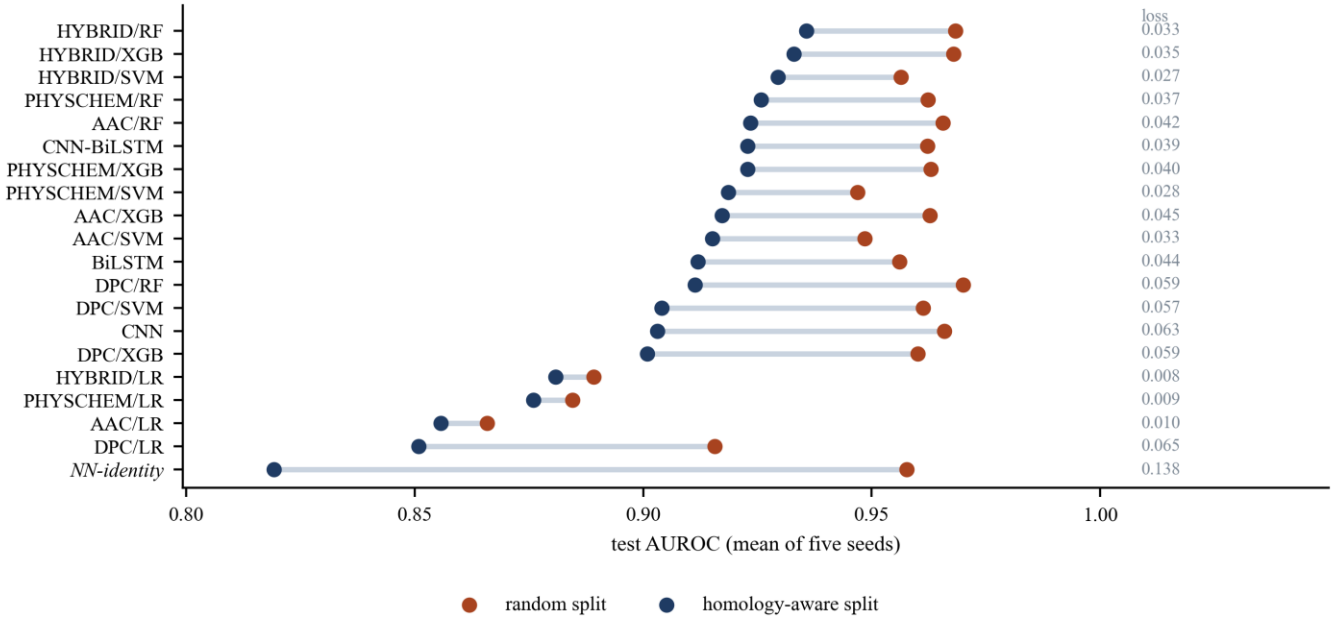


Fig. 2 Every pipeline under both partitioning regimes. The number beside each pair is the accuracy lost when close relatives are removed from the training set.

D. *Accuracy tracks similarity to the training set*

The comparison of regimes hides a continuous relationship, which Figure 3 makes visible. Test peptides from the random splits were binned by their identity to the nearest training peptide. Accuracy was then recomputed inside each bin. For the leading pipeline the area under the curve is 0.992 in the highest identity band, at 0.9 and above. It falls to 0.831 for peptides below 0.6 identity, a gap of 0.161. Panel (b) explains why the headline number sits close to the first figure. Under a random split, 43 per cent of test peptides fall in the highest identity band alone.

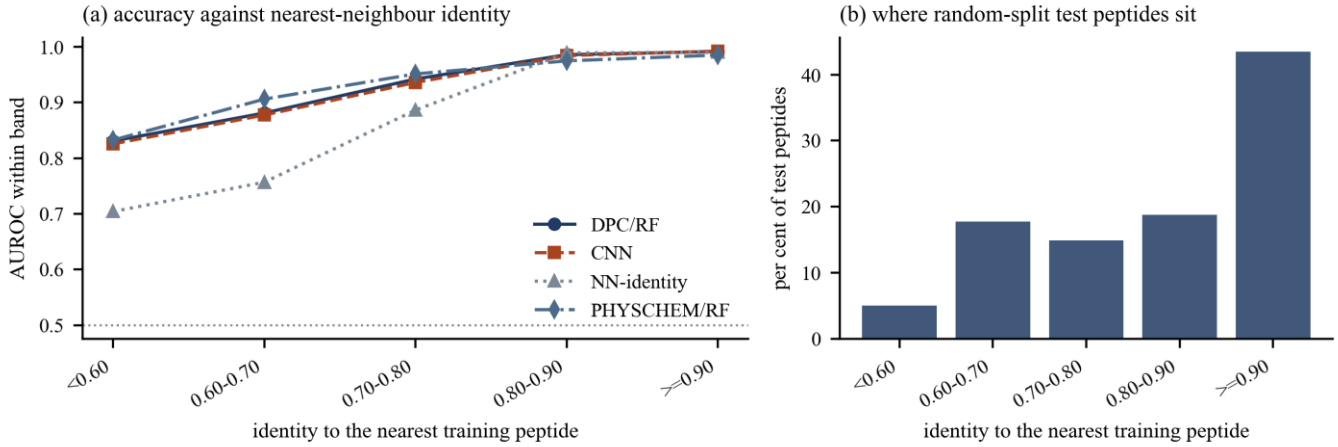


Fig. 3 Accuracy against the identity between each test peptide and its nearest training peptide, pooled over the five random splits.

E. *What separates the two classes*

The biophysical picture agrees with established descriptions of membrane-active peptides. That agreement is a useful check that the benchmark is not an artefact. Table 3 lists the descriptors that separate the classes most strongly. Net charge leads, followed by charge density, isoelectric point and the hydrophobic moment. Antimicrobial peptides here carry a mean net charge of 2.45 at pH 7.4. The negative class carries -0.25. No single descriptor is sufficient on its own, and the strongest reaches an area under the curve of only 0.783.

Table 3. The eight biophysical descriptors that separate the classes most strongly, ranked by rank-biserial effect size. All p values are Holm-corrected.

Descriptor	Antimicrobial	Non-antimicrobial	Effect size	Univariate AUROC	p (Holm)
net charge	2.45	-0.25	0.57	0.783	< 0.0001
charge density	0.09	-0.01	0.52	0.761	< 0.0001
isoelectric point	9.12	7.13	0.46	0.731	< 0.0001
frac hydrophobic	0.46	0.37	0.44	0.718	< 0.0001
aliphatic index	101.88	67.84	0.43	0.713	< 0.0001
hydrophobic moment	0.55	0.43	0.40	0.701	< 0.0001
turn fraction	0.25	0.33	-0.39	0.306	< 0.0001
gravy	0.18	-0.37	0.38	0.691	< 0.0001

IV. Discussion

The central finding is that the partition, not the model, decides what a benchmark of this kind reports. Nineteen trained pipelines all lose accuracy when close relatives are removed from training. They span composition counts, biophysical descriptors, tree ensembles and three neural architectures. The loss averages 0.039, and it is larger than any difference between the pipelines themselves. A control that fits no parameters reaches 0.958 under a random split. That is close to the trained models. Much of what a random split rewards is therefore retrieval rather than generalisation.

The size of the effect deserves a fair statement. The best pipeline still reaches 0.936 under the homology-aware partition, so recognition does not collapse. Cationic charge and amphipathicity are genuine physical correlates of membrane activity, and a model can learn them without memorising

families. Three observations nevertheless stand. Every pipeline loses accuracy, the leading pipeline changes identity, and rank agreement between the regimes is only 0.46. Within a single random split, accuracy ranges from 0.831 to 0.992 depending on how close the nearest training peptide is.

This complements the known negative-selection problem in the field rather than repeating it [12]. That work showed that benchmarks are easier than they look, because the negative class is drawn from a different population. Here both classes were drawn by an identical rule from the same annotation type. That route to inflation is closed by construction. The inflation measured here survives the control, and comes instead from relatedness within each class. The two effects are independent and can act together.

A second practical point concerns how such partitions are built. Holding out whole clusters from a greedy tool is common practice and feels sufficient. It is not. Greedy assignment compares sequences with cluster representatives alone, so relatives can be split across clusters. Here that left 87 per cent of held-out peptides with a relative above the nominal threshold. The transitive closure removes the gap, at the cost of chaining if the threshold is set too low. The two constraints together leave a narrow window of usable thresholds, and that window should be reported.

The dose-response curve gives practical guidance. Accuracy is high for peptides with a close relative in training. It falls steadily as that relative becomes more distant. A predictor is therefore trustworthy in proportion to the similarity between query and training set. That is a statement about applicability domain rather than about the model. Reporting the identity between test peptides and the training set would let readers judge which regime a published number belongs to. Such reporting costs nothing and is easy to automate.

Three limitations should be stated. First, the benchmark uses curated annotation, so it inherits the biases of what has been studied and deposited. Peptide families of medical interest are over-represented, and databases of measured activity would answer a complementary question [30]. Second, activity is treated as binary. The quantity of interest is a minimum inhibitory concentration against a named organism. Third, the neural models here are compact and trained from scratch on a few thousand sequences. Larger pretrained protein language models may transfer better to distant peptides. Testing that claim under a homology-aware partition is the natural next step.

V. Conclusion

This study asked how much reported accuracy in antimicrobial peptide classification depends on sequence homology between training and test data. A benchmark of 6288 mature peptides was assembled from curated annotation, by a rule identical for both classes. Every pair was aligned, and the clustering threshold was calibrated against composition-matched shuffled sequences. Twenty pipelines were then trained under a random partition, and under a partition that removes whole identity clusters.

Three results follow. The peptide set is strongly redundant, and a random split leaves 62 per cent of test peptides with a close relative in training. Every pipeline loses accuracy under the homology-aware partition, by 0.039 on average. That loss exceeds the spread between pipelines. A lookup control that fits nothing reaches 0.958 under the random partition, so similarity alone explains much of the apparent skill.

The practical recommendation is narrow and testable. Papers reporting sequence-based peptide classifiers should partition by identity cluster. They should state the threshold and how it was chosen. They should also report the identity between test sequences and the training set. Doing so lowers the headline number, and makes it mean what readers assume it means.

Acknowledgment

The authors thank the UniProt Consortium for maintaining the curated annotation on which this benchmark is built. The data and the scripts that reproduce every number in this paper are available from the authors on request.

References

[1] C. J. L. Murray et al., "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis," The Lancet, vol. 399, no. 10325, pp. 629-655, 2022.

[2] M. Zasloff, "Antimicrobial peptides of multicellular organisms," Nature, vol. 415, no. 6870, pp. 389-395, 2002.

[3] M. Mahlapuu, J. Hakansson, L. Ringstad, and C. Bjorn, "Antimicrobial peptides: an emerging category of therapeutic agents," Frontiers in Cellular and Infection Microbiology, vol. 6, art. 194, 2016.
[4] C. D. Fjell, J. A. Hiss, R. E. W. Hancock, and G. Schneider, "Designing antimicrobial peptides: form follows function," Nature Reviews Drug Discovery, vol. 11, no. 1, pp. 37-51, 2012.
[5] P. Bhadra, J. Yan, J. Li, S. Fong, and S. W. I. Siu, "AmPEP: sequence-based prediction of antimicrobial peptides using distribution patterns of amino acid properties and random forest," Scientific Reports, vol. 8, art. 1697, 2018.
[6] D. Veltri, U. Kamath, and A. Shehu, "Deep learning improves antimicrobial peptide recognition," Bioinformatics, vol. 34, no. 16, pp. 2740-2747, 2018.
[7] C. Li et al., "AMPlify: attentive deep learning model for discovery of novel antimicrobial peptides effective against WHO priority pathogens," BMC Genomics, vol. 23, art. 77, 2022.
[8] Y. Ma et al., "Identification of antimicrobial peptides from the human gut microbiome using deep learning," Nature Biotechnology, vol. 40, no. 6, pp. 921-931, 2022.
[9] P. Szymczak et al., "Discovering highly potent antimicrobial peptides with deep generative model HydrAMP," Nature Communications, vol. 14, art. 1453, 2023.
[10] M. N. Gabere and W. S. Noble, "Empirical comparison of web-based antimicrobial peptide prediction tools," Bioinformatics, vol. 33, no. 13, pp. 1921-1929, 2017.
[11] J. Xu et al., "Comprehensive assessment of machine learning-based methods for predicting antimicrobial peptides," Briefings in Bioinformatics, vol. 22, no. 5, art. bbab083, 2021.
[12] K. Sidorczuk et al., "Benchmarks in antimicrobial peptide prediction are biased due to the selection of negative data," Briefings in Bioinformatics, vol. 23, no. 5, art. bbac343, 2022.
[13] L. Fu, B. Niu, Z. Zhu, S. Wu, and W. Li, "CD-HIT: accelerated for clustering the next-generation sequencing data," Bioinformatics, vol. 28, no. 23, pp. 3150-3152, 2012.
[14] Y. Park and E. M. Marcotte, "Flaws in evaluation schemes for pair-input computational predictions," Nature Methods, vol. 9, no. 12, pp. 1134-1136, 2012.
[15] S. Whalen, J. Schreiber, W. S. Noble, and K. S. Pollard, "Navigating the pitfalls of applying machine learning in genomics," Nature Reviews Genetics, vol. 23, no. 3, pp. 169-181, 2022.
[16] S. Kapoor and A. Narayanan, "Leakage and the reproducibility crisis in machine-learning-based science," Patterns, vol. 4, no. 9, art. 100804, 2023.
[17] UniProt Consortium, "UniProt: the Universal Protein Knowledgebase in 2025," Nucleic Acids Research, vol. 53, no. D1, pp. D609-D617, 2025.
[18] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," Journal of Molecular Biology, vol. 147, no. 1, pp. 195-197, 1981.
[19] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," Proceedings of the National Academy of Sciences, vol. 89, no. 22, pp. 10915-10919, 1992.
[20] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," Journal of Molecular Biology, vol. 157, no. 1, pp. 105-132, 1982.
[21] D. Eisenberg, R. M. Weiss, and T. C. Terwilliger, "The hydrophobic moment detects periodicity in protein hydrophobicity," Proceedings of the National Academy of Sciences, vol. 81, no. 1, pp. 140-144, 1984.
[22] H. G. Boman, "Antibacterial peptides: basic facts and emerging concepts," Journal of Internal Medicine, vol. 254, no. 3, pp. 197-215, 2003.
[23] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.
[24] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785-794.
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[26] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, art. 6, 2020.
[27] P. J. A. Cock et al., "Biopython: freely available Python tools for computational molecular biology and bioinformatics," Bioinformatics, vol. 25, no. 11, pp. 1422-1423, 2009.
[28] F. Pedregosa et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825-2830, 2011.
[29] A. Paszke et al., "PyTorch: an imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems 32, Vancouver, Canada, 2019, pp. 8024-8035.
[30] M. Pirtskhalava et al., "DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics," Nucleic Acids Research, vol. 49, no. D1, pp. D288-D297, 2021.