XI BIFI International Conference

Artificial Intelligence at the Crossroads of Interdisciplinary Science

January 17th - 19th 2024

bifi24.bifi.es

Instituto Universitario de Investigación Biocomputación y Física de Sistemas Complejos Universidad Zaragoza

Universidad Zaragoza 1542

kampal certest DELL Technologies intel

Book of abstracts

Index

Instituto Universitario de Investigación Biocomputación y Física de Sistemas Complejos
Universidad Zaragoza

Organizers & Sponsors

Organizing Committee

Joaquín Sanz[1,2] Jorge Bertol[1]
David Íñiguez[1,3,4] Beatriz Gómez[4]
María Pilar Latorre Martínez[7] Yolanda Vergara[1]

Scientific Committee

Joaquín Sanz[1,2] Beatriz Herguedas[1,6] Javier García-Nafría[1,6]
David Íñiguez[1,3,4] Beatriz Gómez[4] Sergio Pérez Gaviro[1,2]
Gerardo Sanz[1,5] Ana Cebrián[1,5] Javier López[1,5]
Alberto Aleta[1,2] Alejandro Tejedor[1,2]
Javier Sancho[1,6,7] Yamir Moreno[1,2,8]
María Pilar Latorre Martínez[9]

Conference secretariat

Isabel Nuez[10] María José Pallás[10]
Beatriz Antolí[1] Raúl Mendoza[1]

Sponsors

DELL Technologies
Intel
CERTEST
Kampal Data Solutions

Contact

secretaria@bifi.es
bifi.conferences@bifi.es

---

1. Institute for Biocomputation and Physics of Complex Systems (BIFI). University of Zaragoza.
2. Department of Theoretical Physics, University of Zaragoza.
3. ARAID Foundation.
4. Kampal Data Solutions.
5. Department of statistical Methods, University of Zaragoza.
6. Department of Biochemistry and Molecular and Celular Biology, University of Zaragoza.
7. Aragon Health Research Institute (IIS Aragón), Zaragoza.
8. CENTAI Institute, Turin, Italy
9. Department of Business Management and Organization, University of Zaragoza.
10. Fundación Empresa univiersidad de Zaragoza FEUZ.

Foreword

The 2024 XI BIFI International Conference, themed 'ARTIFICIAL INTELLIGENCE: AT THE CROSSROADS OF INTERDISCIPLINARY SCIENCE,' will unfold from January 17th to 19th, 2024, organized by the Institute for Biocomputation and Physics of Complex Systems at the University of Zaragoza. This edition of the conference highlights the remarkable strides made in the realms of Artificial Intelligence (AI) and Machine Learning (ML) and their increasingly pivotal role in transforming the landscape of scientific research.

In these pages, you will find a curated collection of abstracts that embody the innovative spirit and intellectual rigor of this year's conference. Our gathering brings together a diverse group of leading scientists, each contributing their unique perspectives and expertise in areas where AI and ML methodologies are not merely tools but catalysts for discovery and innovation.

As you explore these abstracts, we invite you to reflect on the extraordinary potential at the intersection of Artificial Intelligence (AI), Machine Learning (ML), and various scientific disciplines. Our collective curiosity, dedication, and collaborative spirit drive us towards a future filled with discoveries and innovation.

The BIFI2024 Organizing Committee

Zaragoza, 2024

Program

Wednesday January 17th

Session 1: 08:30-11:00. Track 1: AI: Methods and applications.
Chairperson: David Íñiguez. ARAID Foundation, BIFI & Kampal

08:30-08:45	REGISTRATION
08:45-09:00	OPENING
09:00-09:40	KEYNOTE SPEAKER: Sofía Teixeira, Universidade de Lisboa Exploring AI and Network Science for Mental Health
09:40-10:00	INTERDISCIPLINARY SESSION. Luis Martín Moreno An Overview of Applications of Deep Learning in Science
10:00-10:20	Juan Luis Durán Batalla Information Retrieval and Keyphrase Extraction with the aid of LLMs
10:20-10:40	Sergio Sabroso Deciphering Pancreatic Ductal Adenocarcinoma progression from AI
10:40-11:00	Douglas V. Laurents ML Programs and NMR Spectroscopy as Partners to Understand Proteins

11:00-11:30	COFFEE BREAK

Session 2: 11:30-13:10. Track 2: AI & Statistical Modeling.
Chairperson: Ana Cebrián, Dept. of Statistical Methods, unizar.

11:30-12:10	KEYNOTE SPEAKER: Paula Gordaliza, Universidad de Navarra Ensuring stability in the assessment of algorithmic fairness
12:10-12:30	INTERDISCIPLINARY SESSION: Luis Mariano Esteban. Modern ML methods: a combination of statistical models & advanced algorithms
12:30-12:50	Rocío Aznar Gimeno. AI and Statistics for Disease Prevention: Practical Cases, Challenges, and Achievements
12:50-13:10	Zeus Gracia-Tabuenca. Predicting depression risk in early adolescence using brain imaging

13:10-15:00	LUNCH

Session 3: 15:00-17:00. Track 2: AI & Statistical Modeling.

Chairperson: Gerardo Sanz, Dept. of Statistical Methods, unizar.

15:00-15:20	Fernando Lahoz Analysis of Random Laser signal in tissues for diagnosis
15:20-15:40	Aurea Grané A compared protocol to improve clustering procedures
15:40-16:00	Manuel Soláns AI for vehicle manufacturing sequence controlling
16:00-16:20	Estela Aguilar Analysis of prediction models to enhance urban bus transportation services through ML techniques
16:20-16:40	Nicolas Béreux Learning a Restricted Boltzmann Machine using biased Monte Carlo sampling
16:40-17:00	Francisco Javier López Analyzing the influence of congestion in user satisfaction: evidence from reviews

17:00-19:10	COFFEE BREAK & POSTER SESSION. (BIFI council: 17:50-19:10; for bifi members only).

- Poster 1. Lorenzo Rosset. Unsupervised hierarchical clustering using the learning dynamics of restricted Boltzmann machines.
- Poster 2. Giovanni Catania. The Copycat Perceptron: Smashing Barriers Through Collective Learning.
- Poster 3. Pablo Pérez Lázaro. Artificial Intelligence techniques and models for cancer diagnosis and risk assessment.
- Poster 4. Marcus Engsig. Enhancing Network Robustness Assessment Through Idle Network Properties.
- Poster 5. Pablo Solán-Fustero. Study on the combination of POD-based ROMs and to augmented Riemann solvers applied to 1D SWE.
- Poster 6. Manuel Gallego. Neural Networks ansatz for solving quantum Long-range models.
- Poster 7. David Muñoz Jordán. Researcher identification and disambiguation across scientific production repositories.

- Poster 8. Marco Fernández Da Silva. Pitch Control Insights from Individual Velocities and Stamina Factors.

- Poster 9. David Luna. Predicting Non-Trivial Lipid Properties Using Neural Networks.

- Poster 10. Alfonso Navas. Inferring effective couplings with Restricted Boltzmann Machines.

- Poster 11. Alberto Aleta. Integrating Social Determinants and Machine Learning to Understand COVID-19 Mortality in Catalonia, Spain.

- Poster 12. Juan Luis Durán. “Knowledge Search Engine”: linking the gap between Academy and Industry.

- Poster 13. Rubén López. Hydrogen Bonding Patterns and Cooperativity in Polyproline II Helical Bundles.

- Poster 14. Oscar Sánchez García. Fluorescence Spectroscopy Combined with Nanoparticles and Machine Learning Analysis for the Diagnosis of Tumoral Pathologies.

- Poster 15. Francisco Javier Falcó. Integration of Machine Learning with Thermal Liquid Biopsy (TLB) using serum samples and patient-specific variables analysis for presurgical diagnosis of malignant ovarian cysts.

- Poster 16. David Polanco. Towards machine learning based diagnosis of Parkinson's disease from two-color fluorescence detection of amyloid aggregates in human biofluids.

- Poster 17. Cristian Segura. Glycine-Rich Polyproline II Helical Bundle Domains Characterized for Improving AI Protein Structure Prediction.

- Poster 18. Pierpaolo Bruscolini. A statistical-physics model for codon usage.

- Poster 19. Carmen Pérez Llantada. Digital language and communication training for EU scientists

- Poster 20. Ignacio Marchante scDVar: a computational tool for modeling inter-individual variation of cell-to-cell transcriptional noise in single-cell RNA-seq.

- Poster 21. Marín Alcalde Comparison between classical linear regression and neural networks for the critical temperature of superconductors.

Thursday January 18th

Session 4: 09:00-11:00. Track 1: AI: Methods and applications.

Chairperson: Luis Martín Moreno, Dept. of Physics of Condensed Matter, unizar.

09:00-09:40	KEYNOTE SPEAKER: Miguel Rocha, Universidade do Minho Supervised and generative deep learning models to predict the activity and to design novel compounds and proteins
09:40-10:00	Miguel A. Vela-Tafalla Genre networks through hyperlinking in science and technology articles written by experts for non-specialized audiences
10:00-10:20	SPONSOR SESSION: Miguel Fajardo, DELL Technologies The AI Boom. A new scenario, a new way of doing things & the role of Dell Tech.
10:20-10:40	Miguel A. Benítez-Castro. Nutcracker, a semi-supervised algorithm for the detection and tackling of online radicalization and extremism
10:40-11:00	Pablo Calvo-Barlés Finding discrete symmetries in data via Machine Learning

11:00-11:30	COFFEE BREAK

Session 5: 11:30-13:10. Track 4: AI applications in Biomedicine.

Chairperson: Beatriz Herguedas,
BIFI & Dept. of Biochemistry and Molecular and Cell Biology, unizar.

11:30-12:10	KEYNOTE SPEAKER: Noelia Ferruz, Barcelona Institute of Molecular Biology Controllable protein design with unsupervised language models
12:10-12:30	Xavier de la Cruz When future is now: the rising contribution of AI to the clinical understanding of genetic variability
12:30-12:50	Carlo Manzo Decoding Microscopic Dynamics through Graph Inductive Knowledge
12:50-13:10	Paulino Gómez-Puertas Molecular dynamics simulations: Applications to the study of macromolecular function and drug design

13:10-15:00	LUNCH

Session 6: 15:00-15:40. Track 2: AI applications in Physics.

Chairperson: Alberto Aleta, BIFI & Dept. of Theoretical Physics, unizar.

15:00-15:40	KEYNOTE SPEAKER: Roger Guimerá, Universitat Rovira i Virgili Bayesian symbolic regression and the learnability of closed-form mathematical models
15:40-16:00	INTERDISCIPLINARY SESSION: Quercus Hernández An overview of deep learning methods for physics simulations
16:00-16:20	Pilar García-Navarro AI-driven computational tools and optimization strategies for hydro-morphodynamic risk prediction & climate change analysis
16:20-16:40	Gabriel Gomila Scanning dielectric microscopy assisted by machine learning
16:40-17:00	Sergio G. Rodrigo Solving differential equations with neural networks

17:00-17:30	COFFEE BREAK

Session 7: 17:30-19:10. Track 2: AI applications in Physics.

Chairperson: Alberto Aleta, BIFI & Dept. of Theoretical Physics, unizar.

17:30-17:50	Pablo F. Garrido A look on human navigation in VR environments by time series clusterization analysis
17:50-18:10	Alexandre Wagemakers Deep Learning-based Analysis of Basins of Attraction
18:10-18:30	Miguel Ruiz-García Out-of-equilibrium machine learning: Dynamical loss functions and catastrophic forgetting
18:30-18:50	Aurélien Decelle. Bipartite Neural Network: effect of non-linearities in the encoding latent variables
18:50-19:10	Gabriel Fernández-Fernández Learning Minimal Representations of Stochastic Processes with Variational Autoencoders

Friday January 19th

Session 8: 09:00-11:00. Track 1: AI: Methods and applications.

Chairperson: Sergio Pérez Gaviro, BIFI & Dept. of Theoretical Physics, unizar.

09:00-09:40	KEYNOTE SPEAKER: Alex Rayón, University of Comillas Emergent abilities in LLMs: myth or truth?
09:40-10:00	Joaquín J. Torres Associative Memory in Dissipative Quantum Neural Networks
10:00-10:20	Rafael Tolosana The AIoD: The European AI On-Demand Platform and Ecosystem
10:20-11:00	KEYNOTE SPEAKER: Beatriz Seoane, Université Paris-Saclay Statistical Physics of Energy-Based generative models

11:00-11:30	COFFEE BREAK

Session 9: 17:30-19:10. Track 4. AI applications in Biomedicine.

Chairperson: Joaquín Sanz, BIFI & Dept. of Theoretical Physics, unizar.

11:30-11:50	INTERDISCIPLINARY SESSION: Ignacio Marchante A role for ML in contemporary Biomedical research.
11:50-12:10	Jonathan Frazer Increasing the diagnostic yield of patient sequencing with proteome-scale probabilistic modelling
12:10-12:30	Helena García-Cebollada Protposer: when machine learning and protein stabilization meet
12:30-12:50	Sonia Hermoso-Durán ML in Thermal Liquid Biopsy of Intracystic Fluid Samples: A New Tool for Presurgical Diagnosis of Pancreatic Cystic Lesions
12:50-13:10	Borja Requena Inferring pointwise diffusion properties of single trajectories with deep learning
13:10-13:30	Javier Orera Application of forward PINN solvers to the modelling of transient blood flow in vessels
13:30-13:45	CLOSING REMARKS

Talks

AI Methods and Applications

Exploring AI and Network Science for Mental Health

Andreia Sofia Teixeira[1]

1. LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, Portugal

Corresponding author: asteixeira@ciencias.ulisboa.pt

With the World Health Organization's alarming estimation that 1 in 4 individuals globally will experience mental or neurological disorders, we are confronted with what some label as an ongoing pandemic. This talk delves into the imperative task of unraveling the intricate mechanisms behind these disorders, emphasizing the important role of AI methods and Network Science.

Beginning with a brief overview of AI methods, we explore their utility as predictive tools for identifying and addressing mental health issues. Navigating the prevalence of mental health disorders, particularly with over 264 million people affected by depression and the stark reality of suicide, claiming nearly 800,000 lives annually, we uncover the potential of network algorithms to comprehend better the structure and complexity of emotional states. We underscore the urgency for an integrative and holistic approach to indicators of mental health disorders.

In conclusion, we explore potential challenges and opportunities in this field, navigating ethical considerations with care. This discussion aims to contribute to a broader understanding of the complexities of mental health, fostering a path toward a more compassionate and comprehensive global approach.

An Overview of Applications of Deep Learning in Science

Luis Martín-Moreno[1,2]

1. Instituto de Nanociencia y Materiales de Aragon (INMA), CSIC-UNIZAR, Spain
2. Departamento de Física de la Materia Condensada, University of Zaragoza, Spain.

*Corresponding author: Luis Martín-Moreno: lmm-*at*-unizar.es*

In this presentation, I will concisely outline the fundamentals of deep learning. Additionally, I will exemplify various problem types that machine learning can address. These examples draw from several bachelor's and master's theses I have supervised over the past few years. The areas covered include noise reduction in experimental data, detection of phase transitions, determination of the number of parameters defining a system, and computation of ground state wavefunctions, among others.

Information Retrieval and Keyphrase Extraction with the aid of Large

Juan Luis Durán Batalla[1], Alejandro Rivero[1,3], Alfonso Tarancón[1,2,3]

1. Kampal Data Solutions.
2. Department of Theoretical Physics, University of Zaragoza.
3. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

*Corresponding author: Juan Luis Durán Batalla: jduran-*at*-kampal.com*

Large Language Models have reshaped the landscape of Natural Language Processing in recent years, especially since the introduction of transformers[2] in 2017. Their versatility spans applications such as classification, question-answering, summarization, information retrieval or topic modelling. Moreover, their widespread accessibility owes much to the availability of open-source pre-trained models. Earlier this year, we introduced a search engine[3] for research production at the Universidad de Zaragoza. In this talk, we will delve into the pivotal role that LLMs assume in two critical facets of the system: the extraction of keywords or keyphrases and the retrieval of information. First, LLMs are combined with classical tools in the field of NLP, such as Levenshtein distances or TF-IDF scoring, to extract keywords from titles and abstracts. Furthermore, LLMs enable fast and multilingual querying, resulting in a straightforward yet effective search engine.

[2] https://arxiv.org/abs/1706.03762

[3] https://buscadordeconocimiento.unizar.es/buscador-conocimiento-search.html

Deciphering Pancreatic Ductal Adenocarcinoma progression from Artificial Intelligence

S. Sabroso-Lasa[1,2,3], E. López de Maturana[1,2], Nannan Xue[1,2], Lola Alonso[1,2], **Núria Malats**[1,2]

1.Genetic and Molecular Epidemiology Group, Spanish National Cancer Research Centre
2. Centro de Investigación Biomédica en Red-Cáncer (CIBERONC), Madrid, Spain.
3. Grupo de Ingeniería y Ciencia de Datos Aplicada, EUPLA.

Corresponding author: Sergio Sabroso-Lasa: ssabroso@cnio.es; Núria Malats: nmalats@cnio.es

Pancreatic Ductal Adenocarcinoma (PDAC), is a low incident tumor with a high mortality rate, mainly due to late detection, and is projected to become the second most mortal type of cancer by 2030. During the last years, a plethora of cancer-specific data were generated by high-throughput biotechnological platforms, facilitating the investigation of cancer biology and classification of patients. On the other hand, this progress is also linked to an increased computational capacity and the constant evolution of AI models that can help us decipher and better understand the mechanisms behind this aggressive disease.

The aim of this presentation is to share an overview of the AI-based work carried out in PDAC's research in the Genetic and Molecular Epidemiology group (GMEG) of the Spanish National Cancer Research Centre (CNIO), focusing on Deep Learning models that allow processing medical image data and predict organ metastasis based on computed tomography (CT) scans. Furthermore, I will present a multimodal data integration of tumor gene expression, mutation, microbiome, radiological images, and pathological slides to stratify PDAC patients based on their prognosis.

Machine Learning Programs and NMR Spectroscopy as Partners to Understand Proteins

Douglas V. Laurents[1]

1. Instituto de Química Física "Blas Cabrera" CSIC, Serrano 119, 28006 Madrid

Corresponding author: Douglas V. Laurents: dlaurents@iqf.csic.es

The machine learning program AlphaFold2 [1] is revolutionizing research in protein Structural Biology as it accurately predicts the 3D structure of two thirds of proteins. These predictions can be used directly as structural models or indirectly to help determine structures experimentally by X-ray crystallography, CryoEM or NMR spectroscopy. Nevertheless, AlphaFold2 and similar machine learning programs such as RoseTTAFold [2] have some weaknesses. They can not provide insight into how proteins fold, nor can they determine protein stability or dynamics. Predicting rare folds or minor alternative conformations accurately remains challenging and AlphaFold2 does not forecast the impact of post translational modifications, mutations or ligand binding. The remaining third of the human proteome which is poorly predicted largely corresponds to intrinsically disordered regions. Essential for regulation and signaling networks, these disordered regions often form biomolecular condensates or amyloids. Fortunately, AlphaFold2's drawbacks are mostly complemented by NMR spectroscopy. This experimental technique provides atomic level information on protein folding and dynamics, biomolecular condensates and amyloids and how they are affected by experimental conditions, post translational modifications, mutations, interactions with small ligands, other proteins and RNA. Together, NMR spectroscopy and AlphaFold2 can collaborate to advance our comprehension of proteins [3].

[1] Jumper J *et. al.* (2021) Nature **596**: 583-589.
[2] Baek M *et. al.* (2021) Science **373**: 871-876.
[3] Laurents DV (2022) Front. Mol. Biosci. **9**: 906437.

Acknowledgement: Este trabajo es parte del proyecto PID2022-137806OB-I00, financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

Supervised and generative deep learning models to predict the activity and to design novel compounds and proteins

Miguel Rocha[1,2]

1. Centre of Biological Engineering - University of Minho
2. CSO - OmniumAI Ltd.

Corresponding author: Miguel Rocha mrocha-at-di.uminho.pt

Artificial Intelligence (AI), and more specifically Machine Learning (ML) models and algorithms, are emerging as important tools to analyze biological sequences and predict distinct phenotypes, as well as to predict the biological activity of chemical compounds. The development of different descriptors to represent DNA/ RNA and protein sequences and compounds has allowed the application of classical ML approaches, while more recently the use of deep learning methods has become dominant in both cases. Here, the surge of distinct vector representations, such as protein or compound embeddings, provided exciting new methods to map compounds/ sequences to functions. Here, we will look at the application of these approaches to predict enzymatic functions from protein sequences or to predict the sweetening power of compounds, as representative examples.

On the other hand, deep generative models have been emerging as important tools to design novel compounds or proteins with desired biological activities or properties. These have been used to create novel chemical compounds and proteins in different areas. Here, we will discuss a case study on the design of new sweeteners, proposing an architecture that combines ML with evolutionary computation to explore the chemical search space.

Genre networks through hyperlinking in science and technology articles written by experts for non-specialized audiences

Miguel A. Vela-Tafalla[1,2]

1. Department of English and German, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

Corresponding author: Miguel A. Vela-Tafalla: mvela@unizar.es

Public engagement with scientific knowledge production and dissemination has brought about a variety of genres and unprecedented relationships among them (Carter-Thomas & Rowley-Jolivet, 2020; Luzón, 2013), creating complex genre assemblages where genres work together ecologically to mediate science communication (Pérez-Llantada & Luzón, 2022; Spinuzzi & Zachry, 2000). This presentation discusses such relationships as established via hyperlinking in articles published in *The Conversation* (www.theconversation.com), a not-for-profit online media outlet where researchers, aided by the editors, discuss recently published research, explain in plain English complex scientific issues, and summarize relevant scholarly debates and controversies for non-specialized audiences. Methodologically, this study explores the possibility of applying corpus linguistics techniques to web text by compiling and directly analysing the source code of the target genre exemplars in a semi-automatic manner, via rule-based web-scraping and HTML parsing. It also explores to what extent it is possible to classify hyperlinks in a semi-automatic way by parsing web addresses, a task typically undertaken by qualitative close reading. Statistical analysis of the data retrieved provides insight into the hypertextual density of these science dissemination texts and discussion of areas for further consideration about *The Conversation* articles. Implications for analysing digital genres are finally discussed.

The Artificial Intelligence Boom. A new scenario, a new way of doing things & the role of Dell Technologies

Miguel Fajardo[1]

1. Director Data Center Sales España – Public Sector - Dell Technologies | Commercial Business

Generative Artificial Intelligence is a technological breakthrough perceived not only in academic and scientific environments but also by the general public. This breakthrough has been made possible by the massive availability of data in digital format, the development by scientists and researchers of algorithms based on deep learning and neural networks, and the emergence of powerful hardware platforms capable of storing and processing these data and algorithms in record time. In this session, we will briefly describe Dell Technologies' contributions to these technologies, including validated designs for data model training and inference environments and some relevant use cases.

Nutcracker, a semi-supervised algorithm for the detection and tackling of online radicalization and extremism

Miguel-Ángel Benítez-Castro[1,2], Juan-Luis Castro-Peña[3], Encarnación Hidalgo-Tenorio[4]

1. Department of English and German Philologies, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Department of Computer Science and Artificial Intelligence, University of Granada.
4. Department of English and German Philologies, University of Granada.

Corresponding author: Miguel-Ángel Benítez-Castro: mbenitez@unizar.es

This presentation is intended to lay out the motivations, rationale and methodological grounds behind the transdisciplinary collaboration between computer scientists and discourse analysts leading to *Nutcracker* (https//nutcracker.ugr.es), a semi-supervised algorithm trained to detect extremist profiles and radicalization in social networking sites. Under the auspices of five national and European research projects starting in 2017 and continuing to date[1], all led by the University of Granada, the fruitful synergies established between computer science and discourse analysis have gradually enabled the training and eventual implementation of a semi-supervised algorithm based upon deep relations aimed at: i) distinguishing extremist from non-extremist profiles; and ii) identifying user networks through similar linguistic realisations of attitude and sentiment (see, e.g. Francisco & Castro, 2020; Francisco, Benítez-Castro, Hidalgo-Tenorio & Castro, 2022). *Nutcracker* relies upon an ontology of key extremist terms identified by the discourse analysts in the team following their corpus-driven examination of a corpus of online Jihadi propaganda, the *JIHAD Corpus* (Dhiab-Hassan, Benítez-Castro & Hidalgo-Tenorio, 2018). Its detection of attitude and sentiment rests upon an ontology derived from Systemic Functional Linguistics' *Appraisal Theory*, and particularly, upon the redefined psychologically-inspired version of the taxonomy proposed in Benítez-Castro & Hidalgo-Tenorio (2019).

Finding discrete symmetries in data via Machine Learning

Pablo Calvo-Barlés[1,2], **Sergio G. Rodrigo**[1,3], **Eduardo Sanchez-Burillo**[4] **and Luis Martín-Moreno**[1,2]

1. Instituto de Nanociencia y Materiales de Aragon (INMA), CSIC-UNIZAR, Spain
2. Departamento de Física de la Materia Condensada, University of Zaragoza, Spain.
3. Departamento de Física Aplicada, University of Zaragoza, Spain.
4. PredictLand S.L., 50001 Zaragoza, Spain

*Corresponding author: Luis Martín-Moreno: lmm-*at*-unizar.es*

Symmetries play an essential role in the understanding of physical theories. A paradigmatic example is the conservation laws arising from continuous symmetries in classical mechanics. Discrete symmetries, such as parity or time reversal, do not depend on any continuous parameter and are key in studying degeneracies in, for instance, the energy spectrum of quantum mechanical systems. Unveiling hidden discrete symmetries is thus of enormous interest in many fields of physics.

In the presence of discrete symmetries in a given system, a group of transformations acts on the physical parameters, ensuring that measurements remain invariant. Calculating the parameters from a given measurement is thus a multivalued problem. Here, we introduce a novel neural network architecture[1], termed Symmetry Seeker Neural Network (SSNN). This network, when provided with experimental or numerical data, can identify all the multivalued solutions associated with symmetries and determine the corresponding representations of the symmetry transformation group.

The SSNN is a powerful tool because it is highly non-trivial to discover symmetries on data sets (especially if the dataset is high-dimensional). The SSNN algorithm is of general applicability. We provide several examples in mathematics, quantum chemistry, and nanophotonics.

[1] P. Calvo-Barlés, S. G. Rodrigo, E. Sánchez-Burillo, and L. Martín-Moreno, Finding discrete symmetry groups via machine learning (2023), arXiv:2307.13457 [physics.comp-ph].

Emergent abilities in LLMs: myth or truth?

Alex Rayón Jerez
Alex.rayon@comillas.edu
University of Comillas

Large language models (LLMs) have shown emergent abilities, demonstrating exceptional performance across diverse tasks for which they were not explicitly trained. Among these, we can find those that require complex reasoning abilities. Only if an LLM has not been trained on a task that it performed we can say that it is showing emergent propertis. Otherwise, the ability must be learned, i.e. through explicit training or in-context learning, in which case it is no longer an ability of the model per-se, and is no longer unpredictable. In other words, the ability is not emergent. In this session, we will understand in which tasks we can identify these emergent capabilities and in which ones not, accordingly, understanding this emergent behavior and its implications.

Associative Memory in Dissipative Quantum Neural Networks

Joaquín J. Torres[1,2]

1. Department of Electromagnetism and Physics of the Matter, University of Granada.
2. Institute "Carlos I" for Theoretical and Computational Physics, University of Granada.

Corresponding author: Joaquín J. Torres: jtorres@onsager.ugr.es

The most prominent case of auto-associative neural networks is the Hopfield model (HM). For several decades this model has been used with great success both in neuroscience to simulate and understand brain processes and in artificial intelligence and machine learning for different tasks including pattern recognition and classification tasks. Recently, quantum versions of the HM have been reported and its memorization and pattern recovery properties have been theoretically studied. The underlying idea is to investigate if these quantum versions of the HM can be used in machine learning for the design of new algorithms that present better performance than their classical versions. In this talk, I will present our recent research concerning the study of the so-called dissipative quantum Hopfield model. The system is constituted of a network of N qubits oscillating at a given frequency Ω and which are coupled via Lindblad jump operators that depend on local fields h_i depending on some given stored patterns. Extensive simulations of the system show pattern-antipattern oscillations of the overlaps with the stored patterns like those reported within a mean-field description, and which are due to metastability originated by the quantum effect driven by the s_x^i qubit operators. Our study shows that such oscillations are stochastic due to the inherent metastability of the pattern attractors induced by the quantum term and disappear in finite systems when one averages over many quantum trajectories. In addition, we have studied the storage capacity of this quantum system and found that the existence of the quantum term of the Hamiltonian has a negative effect on storage capacity, decreasing the overlap with the starting memory pattern for increased values of Ω and the number of stored patterns. However, although the initial pattern destabilizes due to quantum oscillations, other patterns can be retrieved and remain stable for many stored patterns, implying a quantum-dependent nonlinear relationship between the recall process and the number of stored patterns, a fact that could have a potential application in machine learning.

The AIoD: The European Artificial Intelligence On-Demand Platform and Ecosystem

Rafael Tolosana-Calasanz[1]

1. Departmento de Informática e Ingeniería de Sistemas, Universidad de Zaragoza.
2. Instituto Universitario de Investigación en Ingeniería de Aragón, Universidad of Zaragoza.

Corresponding author: Rafael Tolosana-Calasanz: rafaelt@unizar.es

The AIoD platform aims to facilitate the sharing of AI resources, including datasets, algorithms, and software artifacts, as well as computational resources. These assets and computational resources adhere to a unified conceptual schema and governance model, promoting the reproducibility and reuse of AI research findings. The platform is designed to be extensible, allowing the research community to develop and incorporate additional services as needed. AI4Europe is one of the projects, funded under the Horizon Europe programme, that is responsible for the management, development and facilitation of the AI-on-Demand Platform (AIoD).

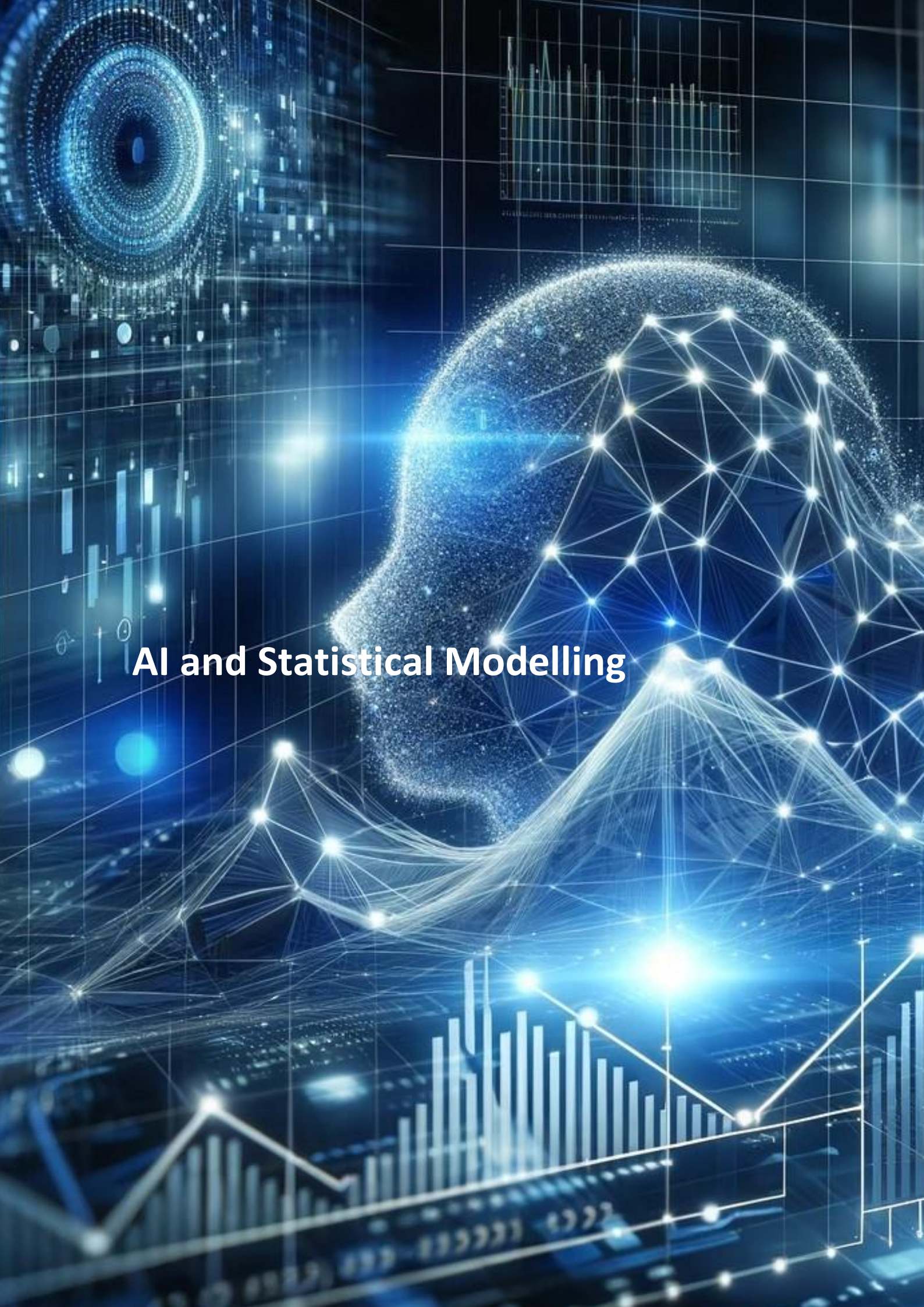
Statistical Physics of Energy-Based generative models

Beatriz Seoane[1]

1. Université Paris-Saclay, CNRS, INRIA Tau team, LISN, 91190, Gif-sur-Yvette, France.

Energy-based models (EBMs) are powerful generative machine learning models that are able to encode the complex distribution of a dataset in the Gibbs-Boltzmann distribution of a model energy function. This means that, if properly trained, they can be used to synthesize new samples that resemble those of the dataset as closely as possible, but also that this energy function can be used to #learn# something about the building mechanisms of the dataset under study.

Indeed, EBMs can be considered a powerful modeling tool for arbitrary data if one were able to map complex energy functions defned in a neural network into spin-interaction Hamiltonians that can be explored using standard statistical physics tools. Such an approach has long been used in physics for inverse Ising problems. The goal now is to extend this approach to more complex energy functions that can encode all higher order correlations in complex data. While this program is very encouraging, training good EBMs is particularly challenging, mainly because they rely on long Monte Carlo sampling processes to estimate the log-likelihood gradient. In my talk, I will present some results on the interpretability of shallow EBMs and discuss how computational statistical physics is a valuable tool for understanding and improving and controlling the training of EBMs.


AI and Statistical Modelling

Ensuring stability in the assessment of algorithmic fairness

Paula Gordaliza[1,2]

1. Institute for Advanced Materials and Mathematics (INAMAT²)
2. Universidad Pública de Navarra.

Corresponding author: Paula Gordaliza: paula.gordaliza-at-unavarra.es

Recent studies highlight the effectiveness of Bayesian methods in assessing algorithm performance, particularly in fairness and bias evaluation. We present Uncertainty Matters, a multi-objective uncertainty-aware algorithmic comparison framework. In fairness-focused scenarios, it models sensitive group confusion matrices using Bayesian updates and facilitates joint comparison of performance (e.g., accuracy) and fairness metrics (e.g., true positive rate parity). Our approach works seamlessly with common evaluation methods like K-fold cross-validation, effectively addressing dependencies among the K posterior metric distributions. The integration of correlated information is carried out through a procedure tailored to the classifier's complexity. Experiments demonstrate that the insights derived from algorithmic comparisons employing the Uncertainty Matters approach are more informative, reliable, and less influenced by particular data partitions.

Modern machine learning methods: a combination of statistical models and advanced algorithms

Luis Mariano Esteban Escaño[1,2]

1. Department of Applied Mathematics, Polytechnic Engineering School of La Almunia, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

Corresponding author: Luis Mariano Esteban Escaño: lmeste@unizar.es

Machine learning algorithms have emerged as a compelling alternative for developing predictive models in recent years. This appeal is particularly attributed to their non-parametric nature, which entails minimal restrictions on their utilization. However, the advancement of these algorithms cannot be solely attributed to this factor; their increased flexibility and adaptability to diverse types of data significantly contribute to their enhanced applicability. The continuous evolution of these models is largely driven by statistical refinements in their training methodologies and code optimizations aligned with computational advancements.

Prominent machine learning models include classification trees, random forest, extreme gradient boosting, ridge regression, LASSO, or elastic net. In this presentation, a comprehensive review of these models will be undertaken, with a focus on elucidating the factors that underlie their advantages over classical techniques.

An additional point of consideration is the inherent dilemma associated with predictions generated by these models. While they generally exhibit heightened accuracy, a concomitant increase in uncertainty is observed.

To illustrate the talk, a prostate cancer database will be used, encompassing 90,000 prostate cancer patients in Aragon who have had at least one PSA measurement in recent years. This database is part of the European project Artificial Intelligence for Health (Ai4health).

AI and Statistics for Disease Prevention: Practical Cases, Challenges, and Achievements

Rocío Aznar Gimeno[1], María Asunción García González[2,3,4], Vega Rodrigálvarez Chamarro[1], Rafael del Hoyo Alonso[1]

1. ITAINNOVA, Aragon Institute of Technology, Spain
2. Instituto de Investigación Sanitaria de Aragón (IIS Aragón), Zaragoza, Spain
3. Instituto Aragonés de Ciencias de la Salud (IACS), Zaragoza, Spain
4. CIBERehd, Zaragoza, Spain

Corresponding author: Rocío Aznar Gimeno; raznar@itainnova.es

Predicting diseases in early stages translates into a significant improvement in the healthcare system, especially for conditions with high prevalence or adverse prognosis. Examples of this include gastric cancer, which has a poor prognosis, that arises from the interaction of various factors, and neurodegenerative diseases, highly prevalent in the adult population. In this context, artificial intelligence (AI) and statistical models play an essential role by handling information and the ability to construct diagnostic and prognostic models.

The talk will address the work and results obtained from public projects focused on the prevention and prediction of these diseases. The development of these projects involves the treatment of various sources of information, such as data and images, and the use of statistical models with the aim of developing IA tools for early diagnosis and decision-making support.

The talk will provide an overview of the process, from information gathering to data treatment, with a particular focus on analysis and modelling. The challenges and limitations of these real problems will also be discussed, along with the achievements made.

Emphasis will be placed on the concept of eXplainable AI, highlighting the importance of model explainability in AI projects, aiming to prevent biases and ensure transparency.

Predicting depression risk in early adolescence using brain imaging

Zeus Gracia-Tabuenca[1,2], Elise B. Barbeau[2], Yu Xia[3], Xiaoqian Chai[2]

1. Department of Statistical Methods, University of Zaragoza.
2. Department of Neurology and Neurosurgery, McGill University.
3. Department of Bioengineering, McGill University

Corresponding author: Zeus Gracia-Tabuenca: zeus@unizar.es

Depression is an incapacitating psychiatric disorder with high prevalence in adolescent populations that is influenced by many risk factors, including family history of depression. It is essential for prevention and intervention strategies to predict who may develop depression before adolescence, when rates increase markedly. Using a large longitudinal sample from 18 hospitals and research centers (final sample: N=2658 participants between 9-10 years at baseline), we applied machine learning methods on a set of neuroimaging features to predict depression risk at the two-year follow-up from the baseline visit. Features include derivatives from structural MRI, diffusion tensor imaging, and task and rest functional MRI. A rigorous cross-validation method of leave-one-site-out was used. Additionally, we tested the prediction models in a high-risk group of participants with parental history of depression (N=625). The results showed all brain features had prediction scores significantly better than chance. When predicting depression onset in the high-risk group, rest fMRI showed the best classification performance, outperforming the other brain features. Results demonstrate that the functional brain imaging can predict the risk of depression in early adolescence better than structural imaging, highlighting the key role of extracting more individual variability in psychopathology compared to focal brain features.

Analysis of Random Laser signal in tissues for diagnosis

F. Lahoz[1], S. de Armas-Rillo[1], B. Abdul-Jalbar[2], D. Luis-Ravelo[3], J. Salas[3], T. González-Hernández[3], J.M. Raya[4]

1. Departamento de Física, IUdEA, Universidad de La Laguna, Spain.
2. Departamento de Matemáticas, Estadística e Investigación Operativa, U. La Laguna, Spain.
3. Departamento de Ciencias Médicas Básicas, ITB, Universidad de La Laguna, Spain.
4. Hospital Universitario de Canarias, Universidad de La Laguna, Spain

Corresponding author: Fernando Lahoz: flahoz@ull.es

Random laser (RL) requires the combination of an optical gain medium with a dispersive environment. The amplified light can be scattered multiple times, resulting in stimulated emission and laser action. RL has been reported in tissues impregnated with a dye solution. The organic dye molecules provide optical gain, while the tissue is a naturally disordered medium that disperses the light. Therefore, the RL emission signal carries structural information of the tissues which has been reported to be useful to differentiate healthy tissues from cancerous ones.

In this investigation we have studied brain samples of transgenic mouse model of Huntington's disease, which is a neurodegenerative disorder characterized by motor and psychiatric symptoms. Brain slices were impregnated with a dye solution. The RL emission data were explored using a multivariate statistical analysis based on principal component analysis and linear discriminant analysis. This statistical analysis allowed us to correctly classify the emission spectra from healthy and from transgenic mice. Moreover, the multivariate statistical analysis was successfully applied to study the effect of an experimental drug on HD transgenic mice. It is expected that artificial intelligence and machine learning methods could improve the analyses of RL data for diagnostic purposes.

Funding: Ministerio de Ciencia e Innovación, Gobierno de España (Grant PID2022-136566NB-I00)

A compared protocol to improve clustering procedures

Aurea Grané[1], Marco Riani[2], Silvia Salini[3]
1. Department of Statistics Universidad Carlos III de Madrid, Spain.
2. Dipartimento di Scienze Economiche e Aziendali, University of Parma.
3. Department of Economics, Management and Quantitative Methods. University of Milan.

Corresponding author: Aurea Grané, agrane@est-econ.uc3m.es

Two widely used machine learning dimensionality reduction techniques are studied, such as t-SNE and UMAP, in the presence of outliers and/or inliers, with the purpose of understanding whether and how they can be used to improve well-known statistical clustering procedures, such as k-means or t-clust.

Artificial Intelligence for vehicle manufacturing sequence controlling.

Manuel Soláns[1]

1. Director IT Supply Chain. Sesé

In the dynamic and demanding automotive industry, efficient production planning and control are crucial for maintaining high quality and timely deliveries. This project delves into the application of Artificial Intelligence (AI) for vehicle manufacturing sequence controlling, specifically utilizing standard classification to monitor sequence deviations in mixed-model production lines.

The automotive industry's pursuit of personalized products at mass-production costs has led to the implementation of order-based sequences for mixed-model assembly lines. However, sequence scrambling, arising from various operational and product-related factors, poses a significant challenge. While Just-in-time (JIT) and fixed production times have been adopted, sequence scrambling remains a persistent issue, disrupting supply chains and production processes.

Early identification of sequence deviations is crucial for timely intervention and sequence stabilization. This project proposes an AI-driven solution to effectively detect and categorize sequence irregularities. Leveraging real-world data from an automotive manufacturer, we train and evaluate a supervised classification model that utilizes components' material, shape, color, and other features to accurately identify sequence deviations.

Our experimental results demonstrate the effectiveness of our approach, achieving up to 95% accuracy in classifying sequence parts. This AI-powered solution provides valuable insights into sequence disruptions and enables proactive corrective actions, enhancing overall production efficiency and reducing downtime.

.

Analysis of prediction models to enhance urban bus transportation services through Machine Learning techniques

Estela Aguilar Margalejo[1,2], Luis Mariano Esteban Escaño[3,4], David Abadía Gallego[2]

1. Master in Modeling and Mathematical research, Statistics and Computing, unizar.
2. ITAINNOVA, Aragon Institute of Technology.
3. Department of Applied Mathematics, Escuela Universitaria Politécnica de La Almunia.
4. Institute for Biocomputation and Physics of Complex Systems, Universidad de Zaragoza.

Corresponding author: Estela Aguilar Margalejo: eaguilar@itainnova.es

Computer vision stands as a fundamental pillar within artificial intelligence, constantly evolving and providing remarkable advancements that impact daily life. In this issue, convolutional Neural Networks (CNNs) provide a powerful tool to process and classify images. In particular, there has been an increase in models with complex architecture and high capacity for processing images from video, YOLO (You Only Look Once) is a convolutional neural network architecture adapted for this type of prediction. This model is currently one of the most utilized for object detection due to its balance between speed and performance.

We have employed YOLO to develop two specific applications within the context of a bus stop. The purpose was to accurately count people waiting and detect wheelchairs and baby strollers within that environment. To ensure robust training, we employed pertinent datasets [1][2]. Results obtained are formally analysed by reasoning in terms of machine learning metrics (such as precision, recall, f1 score) in order to better understand the carried-out object detection and possible improvements to be considered.

The proper implementation of these solutions presents substantial improvements for urban transportation, benefiting both users and the companies responsible for providing this service.

[1] https://storage.googleapis.com/openimages/web/index.html
[2] https://universe.roboflow.com/furnitureselectronics/ultimite_strollers_detection/dataset/1

Learning a Restricted Boltzmann Machine using biased Monte Carlo sampling

Nicolas Béreux[2], Aurélien Decelle[1,2], Cyril Furtlehner[2], Beatriz Seoane[1]

1. Departamento de Física Teórica, Universidad Complutense de Madrid, 28040 Madrid, Spain.
2. Université Paris-Saclay, CNRS, INRIA Tau team, LISN, 91190, Gif-sur-Yvette, France.

Corresponding author: Nicolas Béreux: nicolas.bereux@gmail.com

Restricted Boltzmann Machines are a generative model able to learn any complex distribution from a dataset. Its simple structure makes it particularly useful for interpretability and pattern extraction applications. RBMs, like other energy-based generative models, struggle to describe highly structured data, mainly because their training relies on costly Markov Chain Monte Carlo (MCMC) processes and the cost of sampling multimodal distributions is prohibitive. In particular, we observe that RBMs perform dramatically poorly on artificial low-dimensional clustered datasets.

In our work [2], we investigate a biased sampling method named Tethered Monte Carlo (TMC)[1] to overcome this limitation. This method allows to properly sample such low dimensional datasets in a significantly shorter time, leading to a more accurate likelihood gradient during training, allowing the RBM to accurately learn such datasets. This method can also be used to retrieve the distribution learned by the RBM after training, allowing to assess the quality of the training. However, this method breaks the intra-layer independance of the RBMs which forbids the parallelisation of the MCMC updates, limiting the size of the model we can use.

[1] Martin-Mayor, Seoane and Yllanes, Tethered monte carlo: Managing rugged free- energy landscapes with a helmholtz-potential formalism, Journal of Statistical Physics 144(3), 554 (2011).

[2] Béreux, Decelle, Furtlehner, Seoane, Learning a restricted Boltzmann machine using biased Monte Carlo sampling. SciPost Physics, 14(3), 032 (2023).

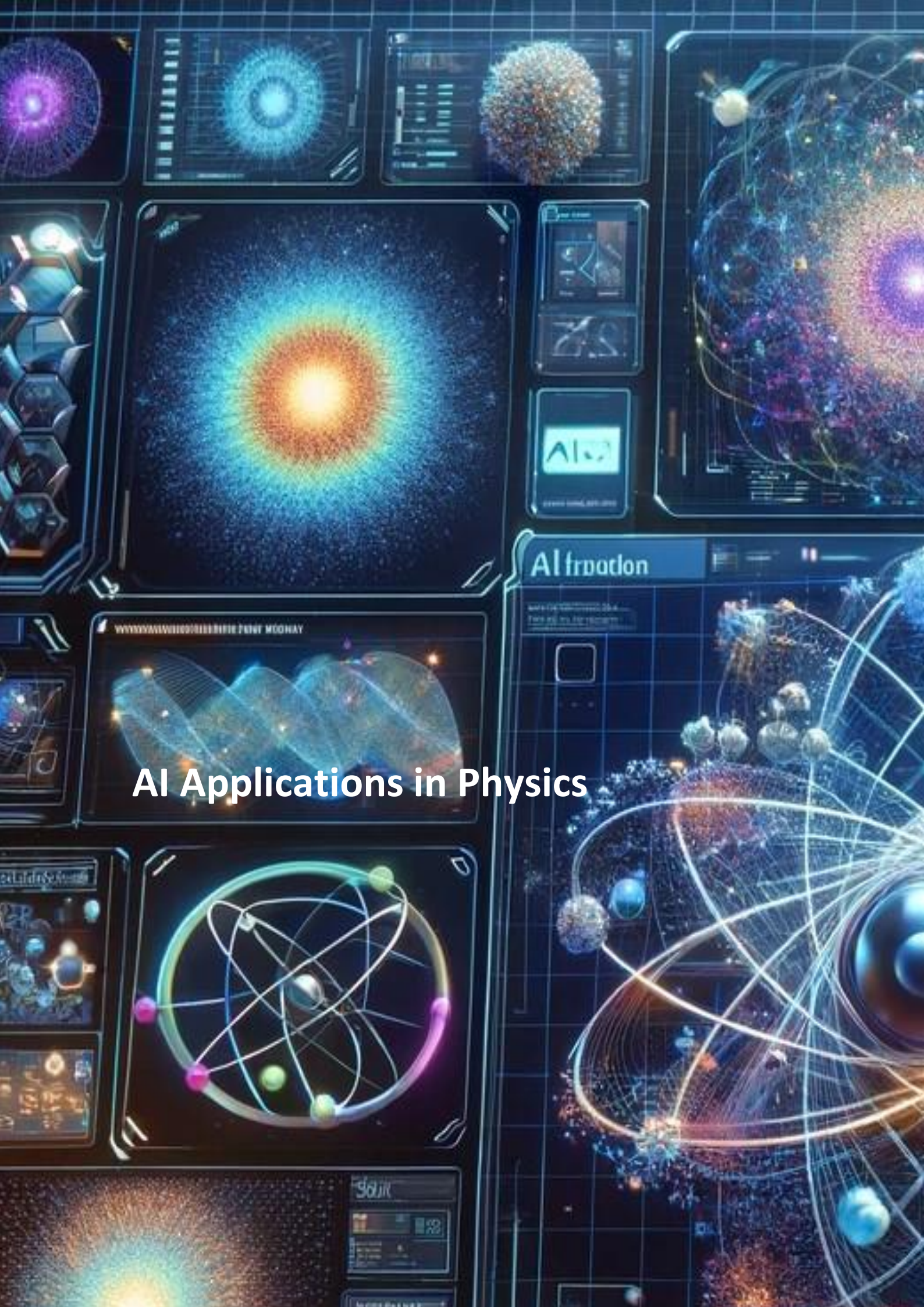
Analyzing the influence of congestion in user satisfaction: evidence from reviews

Alejandro Serrano[1], **F. Javier López**[2,3], Víctor Martínez-de-Albéniz[1]

1. IESE Business School, University of Navarra
2. Department of Statistical Methods, University of Zaragoza
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza

Corresponding author: F. Javier López: javier.lopez@unizar.es

This study investigates the impact of congestion on user satisfaction with touristic services and analyzes factors influencing this relationship. We use Tripadvisor reviews of the 500 most-reviewed attractions globally (2011-2020, ~5M reviews). We measure satisfaction through the rating and complaints about congestion and other questions via text mining of the written reviews. Applying logistic regression models with a large number of fixed effects (BIFE), our findings reveal that the type of attraction significantly shapes sensitivity to congestion. Additionally, staff attentiveness and facilities cleanliness play influential roles. Furthermore, sensitivity is higher for locals in comparison with tourists. Practical implications for stakeholders in the tourism industry are discussed, highlighting the importance of addressing specific aspects such as staff performance and site cleanliness to reduce the impact of long waiting times. Specific actions for local visitors can also be implemented.


AI Applications in Physics

Bayesian symbolic regression and the learnability of closed-form mathematical models

Roger Guimerá[1]

1. Department of chemical engineering, University Rovira I Virgili.

Corresponding author Roger Guimerá roger.guimera-at-urv.cat

Closed-form, interpretable mathematical models have been instrumental for advancing our understanding of the world; with the data revolution, we may now be in a position to uncover new such models for many systems from physics to the social sciences. However, to deal with increasing amounts of data, we need "machine scientists" that are able to extract these models automatically from data. Here, we will introduce the main ideas of Bayesian inference and model selection, and discuss a Bayesian machine scientist, which establishes the plausibility of models using explicit approximations to the exact marginal posterior over models, and establishes its prior expectations about models by learning from a large empirical corpus of mathematical expressions. Within this approach, the space of models is explored using Markov chain Monte Carlo. We will show that this approach uncovers accurate models for synthetic and real data and provides out-of-sample predictions that are more accurate than those of existing approaches and of other nonparametric methods. We will also use this approach to discuss how models are not always learnable from data and that, in some situations, no algorithm will ever be able to learn the correct model from the data alone.

An overview of deep learning methods for physics simulations

Quercus Hernández[1]

1. School of Engineering and Applied Sciences, University of Pennsylvania.

Corresponding author: Quercus Hernández: quercushernandez@gmail.com

A recent interest is observed in the incorporation of already existing scientific knowledge into new deep learning procedures, whose interest is twofold. Indeed, we prefer not to get rid of centuries of scientific knowledge and rely exclusively on powerful machine learning strategies. Existing theories have proved to be useful in the prediction of physical phenomena and are still in the position of helping to produce very accurate predictions. On the other hand, these theories help to keep the consumption of data to a minimum. This is relevant because data are expensive to produce and maintain, so already existing scientific knowledge could alleviate the amount of data needed to produce a successful prediction.

In this talk, a comprehensive overview of several state-of-the-art techniques to simulate dynamical systems will be presented, such as the Physics-Informed Neural Networks (PINN) or Fourier Neural Operators (FNO). The talk will focus on the core concept of inductive bias as a learning prior to incorporate scientific information into the training process. In addition, several specific methodologies for physics problems will be discussed, such as Hamiltonian Neural Networks (HNN) and Structure-Preserving Neural Networks (SPNN).

AI-driven computational tools and optimization strategies for hydro-morphodynamic risk prediction and climate change analysis

García-Navarro P[1,2], Brufau P.[1,2], Martínez-Aranda S.[1,2], Morales-Hernández M.[1,2] Murillo J.[1,2] Navas-Montilla A.[1,2]

1. Fluid Mechanics, University of Zaragoza.
2. TFD Group, I3A, University of Zaragoza.

Corresponding author: P. García-Navarro: pigar@unizar.es

Efficient simulation and optimization tools are crucial for addressing complex hydro-morphodynamic surface processes, which significantly impact natural resources management and risk mitigation in rapidly changing climatic conditions. Over the past decade, the field of geophysical surface processes modelling, including flooding, fertile soil loss, landslides and post-wildfire soil degradation, has significantly grown. This expansion is accelerating, leveraging new computational technologies and addressing emerging challenges, especially those related to climate change and the sustainable use of natural resources. Currently, the integration of real-time satellite imagery with artificial intelligence (AI) enables the analysis of large data volumes for calibration or optimization purposes. This synergy facilitates the development of novel modelling strategies that allow the accurate characterization of different surface flow processes in realistic temporal and spatial scales. Such an approach requires a combination of complex mathematical and numerical modelling, high performance computing (HPC) technologies and advanced optimization techniques powered by AI. Machine learning simulations, in particular, allow for the real-time optimization of models, improving resolution while quantifying uncertainties more effectively. Collectively, these tools and methodologies enable the efficient analysis of complex relationships, aiding decision-making processes in natural systems. Through AI, we can envision more accurate and dynamic models that respond to the ever-evolving landscape of climate change and natural resources management.

Scanning dielectric microscopy assisted by machine learning

M. Checa, [1] R. Millan-Solsona, [2,3] M. Cano, [2] **G. Gomila**[2,3]

1 Center for Nanophase Materials Sciences, Oak Ridge National Laboratory. USA.
2. Department of Electronics and Biomedical Engineering, Universitat de Barcelona. Spain.
3. Nanoscale Bioelectrical Characterization, Institut for Bioengineering of Catalunya. Spain.

Corresponding author: ggomila@ibecbacerlona.eu

Scanning Dielectric Microscopy (SDM) overcomes the limitations of existing nanoscale dielectric imaging methods concerning both data acquisition and quantitative theoretical analysis for its application to highly heterogeneous and nonplanar systems. SDM combines the acquisition of local electric force approach curves at each pixel of the image with its quantitative analysis by using finite element numerical calculations applied to realistic tip-sample geometrical models extracted directly from the measured sample topography. Here, it is demonstrated that the main bottleneck of the technique (the long computation times required to extract the nanoscale dielectric constant maps) can be shortcut by using supervised neural networks, decreasing the computing times from weeks to seconds in a workstation computer. This easy-to-use data-driven approach opens the door for in situ and on-the-fly nanoscale dielectric mapping with scanning dielectric microscopy. Examples of application of the method to the nanoscale mapping of the dielectric properties of eukaryotic cells are provided [1].

[1]. M. Checa, R. Millan-Solsona, A. G. Mares, S. Pujals, G. Gomila, Fast Label-Free Nanoscale Composition Mapping of Eukaryotic Cells Via Scanning Dielectric Force Volume Microscopy and Machine Learning, Small Methods 5, e2100279 (2021).

Solving differential equations with neural networks

Sergio G. Rodrigo [1,2]

1. Departamento de Física Aplicada, Facultad de Ciencias, Universidad de Zaragoza.
2. Instituto de Nanociencia y Materiales de Aragón (INMA), CSIC-Universidad de Zaragoza

Corresponding author: sergut@unizar.es

Artificial Intelligence (AI) encompasses a vast spectrum of computational domains where algorithms are built with the aim to mimic human intelligence, such as image recognition, autonomous driving, and natural language processing. Physics is incorporating these developments in AI and introducing new techniques that could greatly benefit the field.

In recent times, novel AI techniques have emerged to address the resolution of Partial Differential Equations. One of the prominent methods are Physics-Informed Neural Networks (PINNs). PINNs allow solving all types of differential equations, whether ordinary or partial, with one or several variables, single equations or systems of equations. Here, we will present a brief introduction to PINNs, with examples across several fields of physics.

A look on human navigation in VR environments by time series clusterization analysis

Pablo F. Garrido[1,2]

1. Department of Physics, University of Oslo, Norway
2. Center for Lifespan Changes in Brain and Cognition, University of Oslo, Norway.

Corresponding author: p.f.garrido@fys.uio.no

Cognitive training has shown memory improvements in numerous studies, but its usage for better understanding strategies used in human navigation and potential changes with time and training is less explored. To this aim, a unique 3D virtual city model was created in an experimental 10-week cognitive training intervention. While seated on bicycles on bike rolls and wearing virtual reality (VR) masks, participants explored the city several times per week. Additionally, a laptop version of the city was created for home sessions.

A total number of 125 participants, comprising predominantly twin pairs, were included in the experiment. The participants´ position, velocity, and head movement are stored for each trial. Unsupervised Multivariate Time Series Segmentation and Clusterization has been implemented on the VR session data. In addition, a pheromone-like model was created on both the VR and laptop data. The analysis shows a minimum of 5 clusters that are repeated along all the examined sessions and participants. A participant's behaviour during a session is composed by a collection of the detected clusters. The strategy they use depends on their familiarity with the neighbourhood, the distance towards the following target, and on personal variables.

Acknowledgements: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 771355), the UiO:Life Science convergence environment fund, and was developed in collaboration with the Norwegian Twin Registry run by the Norwegian Institute of Public Health.

Deep Learning-based Analysis of Basins of Attraction

Alexandre Wagemakers A[1], David Valle B[1], Miguel A.F. Sanjuán C[1]

1. Nonlinear Dynamics, Chaos and Complex Systems Group, Departamento de Física, Universidad Rey Juan Carlos, Madrid

Corresponding author: Alexandre Wagemakers A: alexandre.wagemakers@urjc.es

This study examines the issue of characterizing the complexity and unpredictability of basins from a variety of dynamical systems. The primary focus is on showcasing the effectiveness of convolutional neural networks (CNNs) in this area. Traditional methods can be computationally demanding when attempting to analyze multiple basins of attraction across various parameters of dynamical systems. This research introduces CNN architectures trained for this task, demonstrating their superior performance compared to conventional techniques. The findings highlight the potential of CNNs to explore different dynamical systems.

Out-of-equilibrium machine learning: Dynamical loss functions and catastrophic forgetting

Miguel Ruiz-García[1,2]

1. *Universidad Complutense de Madrid*
2. Grupo Interdisciplinar de Sistemas Complejos (GISC)

Corresponding author: miguelruizgarcia.eu, miguel.ruiz.garcia@ucm.es

Optimizing a loss function in high-dimensional space lies at the heart of machine learning and other constraint satisfaction problems in physics. The structure of these landscapes and the optimization method used to find solutions can determine the final outcome in both scenarios. We are inspired by the tuning of physical systems which benefit from changes in the loss function landscape [1], to propose a dynamical loss function that helps training and generalization in machine learning [2]. In this talk, we will explore some similarities and differences between constraint satisfaction problems and supervised classification tasks, and show how analogies between the two fields can be exploited to propose new ideas within both realms. Time permitting we will discuss how the model architecture can be modified to transform catastrophic forgetting into positive transfer [3].

[1] Ruiz-Garcia, M., Liu, A. J. and Katifori, E., Tuning and jamming reduced to their minima, Phys. Rev. E 100, 052608 (2019)
[2] Ruiz-Garcia, M., Zhang, G., Schoenholz, S. S., and Liu, A. J., Tilting the playing field: Dynamical loss functions for machine learning, International Conference on Machine Learning, PMLR, 9157 (2021)
[3] Ruiz-Garcia, M., Model architecture can transform catastrophic forgetting into positive transfer, Scientific Reports 12, 10736 (2022)

Bipartite Neural Network: effect of non-linearities in the encoding latent variables

Aurélien Decelle[1], Tony Bonnaire[2], Giovanni Catania[1], Beatriz Seoane[3]

1. Department of Theoretical Physics, University Complutense of Madrid (UCM).
2. Physics laboratory of École Normale Supérieux Paris, Sorbonne University.
3. LISN Tau team, Université Paris-Saclay

Corresponding author: ***Aurélien Decelle*** *adecelle@ucm.es*

In recent decades, generative neural networks have been extensively explored as systems that can store different global configurations. The Hopfield model is a paradigmatic example of this mechanism. In this presentation, we will focus on generative bipartite neural networks described by the Gibbs-Boltzmann distribution. Depending on its parameters, the architecture of this network can either adapt to the Hopfield model or evolve into more complex models such as Restricted Boltzmann Machines.

In this discussion, we will discuss how the properties of the stochastic variables significantly affect the memorization capabilities and the phase diagram of the network. We will also explore strategies to bypass the spin-glass phase and thus effectively retrieve the patterns embedded in the models without a prior knowledge of them.

Learning Minimal Representations of Stochastic Processes with Variational Autoencoders

Gabriel Fernández-Fernández,[1] Carlo Manzo,[2,3] Maciej Lewenstein,[1,4] Alexandre Dauphin,[1] and Gorka Muñoz-Gil [5]

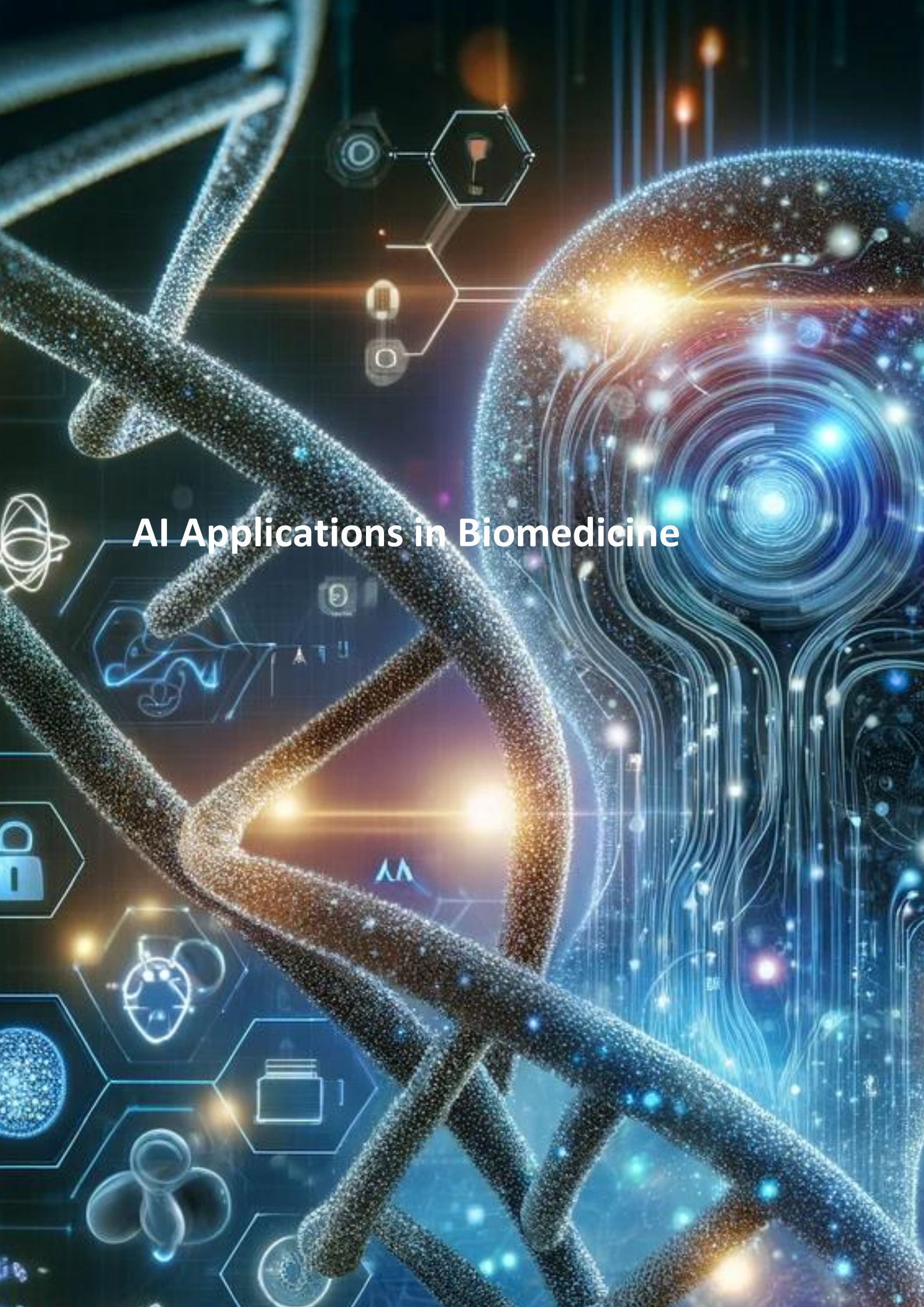
1. ICFO – Institut de Ciències Fotòniques,
The Barcelona Institute of Science and Technology,
Av. Carl Friedrich Gauss 3, 08860 Castelldefels (Barcelona), Spain
2. Facultat de Ciències, Tecnologia i Enginyeries,
Universitat de Vic – Universitat Central de Catalunya (Uvic-UCC),
C. de la Laura, 13, 08500 Vic, Spain
3. Institut de Recerca i Innovació en Ciències de la Vida i de la Salut
a la Catalunya Central (IRIS-CC), 08500 Vic, Barcelona, Spain
4. ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain
5. Institute for Theoretical Physics, University of Innsbruck,
Technikerstr. 21a, A-6020 Innsbruck, Austria

Corresponding author:
Gabriel Fernández-Fernández: gabriel.fernandez@icfo.eu

Machine learning (ML) is an increasingly important tool in science, and understanding why and how the machines predicts is essential. The field of interpretable ML has developed approaches that may provide valuable insights into the underlying data patterns and, thus, aid in the development of new hypotheses. Moreover, these approaches can benefit from the controlled ground provided by the models of statistical physics, in particular, stochastic processes describing the molecular motion from single-particle tracking data.

Here, we aim to use interpretable ML methods to extract, unsupervisedly, the relevant physical parameters of given sets of stochastic trajectories. For this task, we extend the β-variational autoencoder (β-VAE) to account for the typical properties found in diffusion. To show the validity of this approach, I will describe the application of β-VAE to simulated data from stochastic processes such as Brownian motion, fractional Brownian motion, and scaled Brownian motion. Our analysis of the β-VAE reveals its ability to help identify relevant data parameters, and generate new trajectories with the learned statistics. Our future perspectives include the application of this technique to other stochastic processes, as well as experimental applications where one can extract the minimal representation of trajectories and establish connections with existing theories.

See the preprint at arXiv:2307.1https://doi.org/10.48550/arXiv.2307.116081608. To be published soon.


AI Applications in Biomedicine

Controllable protein design with unsupervised language models

Noelia Ferruz[1]

1. Department of Structural and Molecular Biology, Molecular Biology Institute of Barcelona (IBMB-CSIC), Barcelona Science Park, 08028 Barcelona, Spain.

Corresponding author: noelia.ferruz@ibmb.csic.es

Artificial Intelligence (AI) methods are emerging as powerful tools in fields such as Natural Language Processing (NLP) and Computer Vision (CV), impacting the tools and applications we use in our daily lives. Language models have recently shown incredible performance at understanding and generating human text, producing text often indistinguishable from that written by humans. Inspired by these recent advances, we trained a language model, ProtGTP2, which effectively learned the protein language and generated sequences in unexplored regions of the protein space. A desirable critical feature in protein design is having control over the design process, i.e., designing proteins with specific properties. For this reason, we trained ZymCTRL, a model trained on enzyme sequences and their associated Enzymatic Commission (EC) numbers. ZymCTRL generates enzymes upon user-defined specific catalytic reactions, which show natural-like catalytic activities in wet lab experiments. Lastly, we have trained REXzyme, a translation machine capable of designing enzyme sequences for user-defined chemical reactions.

When future is now: the rising contribution of AI to the clinical understanding of genetic variability

Xavier de la Cruz[1,2]

1. Vall d'Hebron Institute of Research (VHIR).
2. Institució Catalana de Recerca i Estudis Avançats (ICREA).

Corresponding authorXavier de la Cruz: xavier.delacruz@vhir.org

After the Human Genome Project raised high expectations for the influence of genomics on medicine, Precision Medicine (PM) emerged with the promise of customizing treatments based on individual genetic profiles. However, it soon became clear that there were significant scientific and technical challenges to fulfilling this promise. Among these challenges is the 'variant interpretation problem (VIP)', a particular case of the central genotype-to-phenotype problem in Biology. The VIP involves determining whether a particular genetic variant is pathogenic or not. The scientific complexity of this problem has slowed PM progress and led to considerable frustration in the biomedical field.

This challenge has accelerated the development of computational methods, particularly those using Artificial Intelligence (AI), to address the VIP. These techniques are beneficial because they tackle the VIP indirectly, without needing to initially resolve the deep scientific questions involved. In my presentation, I'll share examples from our research and others' to show how our approach to this issue has advanced and how our use of AI has adapted over time. Specifically, I'll discuss the shift in focus not just on performance but also on the interpretability of results, aligning with concerns raised by the scientific community and the European Union regarding the role of Artificial Intelligence in high-risk applications.

Decoding Microscopic Dynamics through Graph Inductive Knowledge

Jesús Pineda[1], Benjamin Midtvedt[1], Harshith Bachimanchi[1], Giovanni Volpe[1], **Carlo Manzo**[2,3].

1. Department of Physics, University of Gothenburg.
2. Universitat de Vic – Universitat Central de Catalunya (UVic-UCC).
3. Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC).

Corresponding author: Carlo Manzo, carlo.manzo@uvic.cat

The characterization of dynamic processes in living systems yields essential insights about life processes in health and diseases. Significant advancements in optical microscopy enable us to explore the motion of cells, organelles, and individual molecules with unprecedented detail. Nevertheless, analyzing the dynamic processes in complex and crowded environments remains a challenge.

This work introduces MAGIK (Motion Analysis through Graph Inductive Knowledge), a geometric deep-learning framework tailored for decoding biological dynamics in time-lapse microscopy. MAGIK models movement and interactions through a directed graph, where nodes represent detections and edges connect spatiotemporally close nodes. The framework employs a graph neural network to process the graph and modulate the association strength between nodes via a learnable local receptive field and a gated self-attention mechanism.

MAGIK demonstrates remarkable versatility and excels in multiple tasks, from linking coordinates into trajectories to inferring local and global dynamic properties. MAGIK's flexibility and reliability are demonstrated through applications to real and simulated data across a broad range of biological experiments, underscoring its potential to significantly contribute to our understanding of dynamic processes in living systems.

Pineda, J. *et al.* **Geometric deep learning reveals the spatiotemporal features of microscopic motion**. *Nat Mach Intell* **5**, 71–82 (2023)
https://doi.org/10.1038/s42256-022-00595-0

Molecular dynamics simulations. Applications to the study of macromolecular function and drug design

David Ros-Pardo[1], Íñigo Marcos-Alcalde[1] and **Paulino Gómez-Puertas**[1]

1. Centro de Biología Molecular Severo Ochoa (CBMSO, CSIC-UAM). 28049 Madrid, Spain.
Corresponding author: Paulino Gómez-Puertas <pagomez@cbm.csic.es>

In recent years, three different scales of simulation have been used in the laboratory: a subatomic scale (QM/MM simulations), where variations in the electronic orbitals of the atoms of interest are taken into account; an atomic scale (molecular dynamics), where atoms are treated as indivisible entities and interactions between them are simulated; and a coarse-grained scale (coarse-grained molecular dynamics), where the basis of the simulation is groups of atoms with common physicochemical properties. The differences between these scales are the detail of the model, from subatomic structures to large molecular complexes, and the simulation time, from a few nanoseconds at the most detailed scale to several microseconds at the coarsest scale.
Examples of each of these systems in the study of protein function and their use in drug design (peptides or small molecules) will be discussed. Examples of the use of the AlphaFold artificial intelligence system for protein structure prediction will also be discussed.

A role for ML in contemporary Biomedical research

Ignacio Marchante Hueso[1,2]

1. Department of Theoretical Physics, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, BIFI, unizar.

In recent years, the convergence of technological advancements in experimental techniques and the ever-increasing availability of computing resources has inaugurated a new era in Biological and Medical Research. This era is characterized by an expanding ability to produce and analyze quantitative data with increasing resolution, throughput, and volume across various fields, ranging from genomics and systems biology to structural biology and biophysics. Consequently, machine learning (ML) techniques emerge as a powerful toolbox to provide interesting insights and useful predictive models for novel biomedical questions of unprecedented depth.

In this talk, we will provide a broad overview of some of the main applications of ML in contemporary biomedical research, showcasing recently developed examples of (both supervised and unsupervised) algorithms and models. More precisely, we will discuss three fields where these techniques are exerting a profound impact. First, we will delve into the usage of ML in computational genomics, encompassing applications involving various -omics and multi-omics types of data. Second, we will overview the revolutionary role that ML is playing in the study of protein structure and function, providing insight into questions related to key biophysical processes such as protein folding, interactions, and stability, and enabling researchers to predict protein function or even design novel proteins *de novo*. Lastly, we will explore the application of ML techniques to characterize stochastic processes in the dynamics of biomolecular systems.

Increasing the diagnostic yield of patient sequencing with proteome-scale probabilistic modelling

Jonathan Frazer[1]

1. Centre for Genomic Regulation CRG, Barcelona.

Corresponding author: Jonathan Frazer jonathan.frazer-at-crg.eu

Identifying causal mutations accelerates genetic disease diagnosis, and therapeutic development. While computational prediction methods are increasingly successful at predicting disease-causing variants in known disease genes, they do not generalize well to other genes as the scores are not calibrated across the proteome. To address this, we developed a deep generative whole-proteome model, popEVE, that combines evolutionary information with population sequence data and achieves state-of-the-art performance at distinguishing patients with severe developmental disorders from healthy individuals. popEVE provides compelling evidence for genetic diagnoses even in exceptionally rare single-patient disorders where conventional techniques relying on repeated observations are not applicable. Analysing a cohort of 31k patients we find evidence of over 100 novel genetic disorders.

Protposer: when machine learning and protein stabilization meet

Helena García-Cebollada[1,2], Alfonso López[1,2], Javier Sancho[1,2,3]

1. Department of Biochemistry, Molecular and Cell Biology, Faculty of Science, University of Zaragoza, 50009 Zaragoza, Spain.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Aragon Health Research Institute (IIS Aragón), 50009 Zaragoza, Spain

Corresponding author: Helena García-Cebollada: hgarcia-at-*unizar.es*

Protein-based products are now present in everyday life. However, stability is a key feature for their preservation and usage in convenient extreme conditions. For this purpose, we have developed Protposer, a web server that combines an algorithm for mutation proposal based on rational approaches to protein stabilization with a machine learning model for mutation evaluation, so that a user without deep knowledge of structural biology can get probably stabilizing mutants of their protein of interest. After training and testing the logistic regression model with thermodynamic data from ProTherm database, its positive predictive value (PPV) was better than current stability predictors. When combined with the proposal algorithm, its performance improved even more, outperforming Rosetta, FoldX and PoPMuSiC, and obtaining a total stabilization over 8 kcal/mol for the Nostoc apoflavodoxin with PDB ID 1FTG. Currently, an improved Premium version is under development. The server is available free for academic purposes and under license for commercial purposes in http://webapps.bifi.es/the-protposer.

Machine Learning (ML) in Thermal Liquid Biopsy (TLB) of Intracystic Fluid Samples (IFS): A New Tool for Presurgical Diagnosis of Pancreatic Cystic Lesions (PCL)

Hermoso-Durán, Sonia[1,2,3], Nicolas Fraunhoffer[4], M. Ortiz de Solórzano Reig[5], J. Millastre Bocos[5], G. García Rayado[5], Ángel Lanas Arbeloa[2,5], Adrián Velázquez-Campoy[1,2,3,6], Olga Abian[1,2,3,6]

1. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Spain.
2. Aragón Health Research Institute (IIS-Aragón), Zaragoza, Spain.
3. CIBERehd, Spain.
4. Centre de Recherche en Cancé-rologie de Marseille (CRCM), France.
5. Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain.
6. Department of Biochemistry and Molecular and Cell Biology, University of Zaragoza, Spain.

Corresponding author: Sonia Hermoso Durán: shermosod@gmail.com

Mucinous PCLs present a challenge in clinical diagnosis due to their high malignant potential. To address this, we introduce the use of TLB to analyze IFS, developing a classification model, taking advantage of ML algorithms, to improve the diagnostic accuracy of these neoplasms.

In this study, IFS were analyzed using TLB. Our methodology uses temperature pair analysis, employing cross-validation techniques and Lasso-type penalties, for the selection of significant variables. Model training covered 70% of the data set, with a 100x randomization process. We designed a first model, TLB1, to discriminate between mucinous and non-mucinous cysts, while a TLB2, to differentiate malignant from non-malignant mucinous cysts.

We utilized 35 samples for TLB1 and 17 for TLB2. The analysis revealed differences in the median values of both models (p-value<0.001). The area under the curve was 0.89 and 0.99, respectively, and both models achieved a negative predictive value of 100%.

The integration of ML algorithms into the analysis of IFS thermogram curves represents a significant advancement, improving its differentiation. However, additional research with larger sample sizes is imperative, to confirm our findings and to pave the way for more precise and reliable diagnostic methods in potentially malignant PCL.

Inferring pointwise diffusion properties of single trajectories with deep learning

Borja Requena[1], Sergi Masó-Orriols[2,3], Joan Bertran[2,3], Maciej Lewenstein[1,4], Carlo Manzo[2,3], Gorka Muñoz-Gil[5]

1. ICFO – Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona), Spain.
2. Facultat de Ciències, Tecnologia I Enginyeries, Universitat de Vic – Universitat Central de Catalunya (UVic-UCC), Vic, Spain.
3. Institut de Recerca i Innovació en Ciències de la Vida i de la Salut a la Catalunya Central (IRIS-CC), Vic, Barcelona, Spain
4. ICREA, Pg. Lluís Companys 23, Barcelona, Spain
5. Institute for Theoretical Physics, University of Innsbruck, Innsbruck, Austria

Corresponding author: Borja Requena: borja.requena@icfo.eu

To characterize the mechanisms governing the diffusion of particles in biological scenarios, it is essential to accurately determine their diffusive properties. To do so, we propose a machine-learning method to characterize diffusion processes with time-dependent properties at the experimental time resolution. Our approach operates at the single-trajectory level predicting the properties of interest, such as the diffusion coefficient or the anomalous diffusion exponent, at every time step of the trajectory. In this way, changes in the diffusive properties occurring along the trajectory emerge naturally in the prediction and thus allow the characterization without any prior knowledge or assumption about the system. We first benchmark the method on synthetic trajectories simulated under several conditions. We show that our approach can successfully characterize both abrupt and continuous changes in the diffusion coefficient or the anomalous diffusion exponent. Finally, we leverage the method to analyze experiments of single-molecule diffusion of two membrane proteins in living cells: the pathogen-recognition receptor DC-SIGN and the integrin α5β1. The analysis allows us to characterize physical parameters and diffusive states with unprecedented accuracy, shedding new light on the underlying mechanisms.

Application of forward PINN solvers to the modelling of transient blood flow in vessels

J. Orera, Julia Ramírez, Juan Mairal, Pilar García-Navarro, Javier Murillo

I3A, Universidad de Zaragoza, C/Mariano Esquillor s/n, 50018, Zaragoza, Spain

Corresponding author: jorera@unizar.es

Blood flow dynamics can be modeled using one-dimensional approaches [1], leading to a hyperbolic system of equations with geometrical source terms. Applications of relevance are associated to the spatial variation of mechanical and geometrical properties describing the blood-wall fluid-structure interaction, which requires the use of advanced numerical solvers [2]. If source terms are deficiently tackled, spurious oscillations may be generated in the numerical solution leading to prediction failure [3].

In this work, we leverage the use of forward Physics-Informed Neural Network (PINN) solvers [4] in combination with the aforementioned numerical methods, to provide a description of the propagation of nonlinear waves inside vessels [5], using a reduced model of the system of equations defined by mass and momentum conservation. This hybrid approach is tested against traditional *vanilla* PINNs, where the reduced model is incorporated via automatic differentiation.

Bibliography:

[1] Jordi Alastruey, KH Parker, J Peiró, SJ Sherwin, "Lumped parameter outflow models for 1-D blood flow simulations: effect on pulse waves and parameter estimation," Communications in Computational Physics, Vol. 4, No. 2, pp. 317–336, 2008.

[2] Lucas O. Müller, Pablo J. Blanco, "A high order approximation of hyperbolic conservation laws in networks: Application to one-dimensional blood flow," Journal of Computational Physics, Vol. 300, pp. 423-437, 2015.

[3] J. Murillo, A. Navas-Montilla, P. García-Navarro, "Formulation of exactly balanced solvers for blood flow in elastic vessels and their application to collapsed states," Computers and Fluids, Vol. 186, pp. 74-98, 2019.

[4] Maziar Raissi, Paris Perdikaris, George E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, Vol. 378, pp. 686-707, 2019.

[5] J. Murillo, P. García-Navarro, "A Roe type energy balanced solver for 1D arterial blood flow and transport," Computers and Fluids, Vol. 117, pp. 149-167, 2015.

Posters

Poster 1

Unsupervised hierarchical clustering using the learning dynamics of restricted Boltzmann machines

Lorenzo Rosset[1,2], Aurélien Decelle[2,3], Beatriz Seoane[2,3]

1. Laboratoire de Physique de l'Ecole Normale Superieure, ENS, Universite PSL, CNRS, Sorbonne Universite, Universite de Paris, Paris, France.
2. Departamento de Física Teórica, Universidad Complutense de Madrid, 28040 Madrid, Spain.
3. Université Paris-Saclay, CNRS, INRIA Tau team, LISN, 91190 Gif-sur-Yvette, France.

Corresponding author: Lorenzo Rosset: lorenzo.rosset@phys.ens.fr

Data sets in the real world are often complex and to some degree hierarchical, with groups and subgroups of data sharing common characteristics at different levels of abstraction. Understanding and uncovering the hidden structure of these data sets is an important task that has many practical applications. To address this challenge, I will present a general method for building relational data trees by exploiting the learning dynamics of the restricted Boltzmann machine. Our method is based on the mean-field approach, derived from the Plefka expansion, and developed in the context of disordered systems. It is designed to be easily interpretable. We tested our method in an artificially created hierarchical data set and on several different real-world data sets (images of digits, mutations in the human genome, and a homologous families of proteins). The method is able to automatically identify the hierarchical structure of the data. This could be useful in the study of homologous protein sequences, where the relationships between proteins are critical for understanding their function and evolution.

Poster 2

The Copycat Perceptron: Smashing Barriers Through Collective Learning

Giovanni Catania[1], Aurélien Decelle[1,2], Beatriz Seoane[2]

1. Departamento de Física Teórica, Universidad Complutense de Madrid, 28040 Madrid, Spain
2. Université Paris-Saclay, CNRS, INRIA Tau team, LISN, 91190 Gif-sur-Yvette, France

Corresponding author: Giovanni Catania: gcatania@ucm.es

We characterize the equilibrium properties of a model of coupled binary perceptrons in the teacher-student scenario, with an explicit ferromagnetic coupling favoring configurations where students have similar weights, to mimic an simple collaborating protocol between different students. Using analytic techniques commonly used in spin-glass theories, we characterize the equilibrium properties of such a model in the presence of a thermal noise. In the nonzero temperature regime, we find that the coupling of replicas bends the phase diagram towards smaller values of the fraction of samples provided to each student: This suggests that the free energy landscape gets smoother around the solution with perfect generalization (i.e the teacher), allowing thermal algorithms such as Simulated Annealing to easily reach the teacher solution and avoid getting trapped in metastable states as it happens in the unreplicated case in a specific regime of the phase diagram. On top of provididing additional analytic and numerical evidence for the recently conjectured Bayes-optimal property of Replicated Simulated Annealing (RSA), these results also suggest that multiple students working together are able to learn the same rule faster and with fewer examples, a property that could be exploited in the context of cooperative and federated learning.

Poster 3

Artificial Intelligence techniques and models for cancer diagnosis and risk assessment

Pablo Pérez Lázaro[1,2], Rocío Aznar Gimeno[2], Ana C. Cebrián Guajardo[3,4], Vega Rodrigálvarez Chamarro[2], Rafael del Hoyo Alonso[2]

1. Master in Quantitative Biotechnology Student, University of Zaragoza.
2. ITAINNOVA, Aragon Institute of Technology.
3. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
4. Department of Statistical methods, University of Zaragoza.

Corresponding author: Pablo Pérez Lázaro: plazaro@itainnova.es

This research project focuses on predicting the probability of presenting early-stage colorectal cancer (CRC) using AI algorithms. For this purpose, a comprehensive dataset with demographic and clinical details of approximately 25,000 patients who underwent colonoscopies in Sector III of Zaragoza between 2015 and 2019 is available.

The project methodology involves data handling and preparation to ensure dataset cleanliness, completeness and consistency, as well as data analysis. For instance, K-modes clustering was used to identify patient groups based on their comorbidities, allowing the categorization of patients based on common patterns.

Prior to model training, different statistical techniques and AI algorithms will be explored to select the most suitable one. Additionally, explanaible AI (XAI) techniques will be investigated to provide interpretability of the models, offering a deeper explanation of how the model arrives at a particular result. This critical aspect ensures the practical application of predictive models, providing clinicians with valuable information for personalized interventions and follow-ups.

The ultimate goal of the project is to develop a decision support tool that integrates these AI models to assess the risk of CRC based on patient information.

This poster will present an overview of the methodology, highlight current results and outline upcoming research steps.

Poster 4

Enhancing Network Robustness Assessment Through Idle Network Properties

Marcus Engsig[1], **Alejandro Tejedor**[2,3], Yamir Moreno[2,3,4]

1. Directed Energy Research Centre, Technology Innovation Institute.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Department of Theoretical Physics, University of Zaragoza.
4. CENTAI Institute

Corresponding author: Alejandro Tejedor: alej.tejedor@gmail.com

Network robustness, essential for sustained system functionality amid failures or targeted attacks, has traditionally been evaluated based on the connectivity of nodes resilient to attacks. This research introduces a novel perspective by emphasizing the significance of idle network properties—connectivity of nodes affected by attacks—in assessing overall network robustness. Our study reveals that integrating idle network properties provides important insights, improving the accuracy of robustness assessment and enhancing the ability to generalize models for estimating robustness in unforeseen attack scenarios.

Methodologically, we frame the problem as a regression task, comparing the estimation accuracy of neural networks trained solely on active indicators against those incorporating both active and idle indicators. Our results suggest that incorporating idle network information enables more accurate estimations tailored to specific network topologies and attacks. Furthermore, it equips models with the ability to navigate in-sample and out-sample variability, marking a conceptual leap in assessing network robustness. Thus, incorporating idle information in assessing network robustness has the potential to guide the development of novel strategies for enhancing network resilience and offers insights into the design and reconfiguration of existing networks to maximize their structural robustness.

Poster 5

Study on the combination of POD-based ROMs and to augmented Riemann solvers applied to 1D SWE

Pablo Solán-Fustero[1], José Luis Gracia[2], Adrián Navas-Montilla[1], Pilar García-Navarro[1]

1. I3A and Fluid Mechanics Department, University of Zaragoza.
2. IUMA and Applied Mathematics Department, University of Zaragoza.

Corresponding author: psolfus@unizar.es

The shallow water equations (SWE) are a widely used to mathematically model surface flows such as rivers, coastal flows, atmospheric events, or lava flows. Since the SWE do not have an analytical solution, they have to be solved numerically. The work of Fluid Dynamic Technologies Group (I3A) is focused on the development of software to compute realistic problems when using the SWE. The complex and variable nature of realistic scenarios requires the use of advanced numerical schemes, which include augmented Riemann solvers, and this involves high computational costs. With the aim of preserving predictability, it is very interesting to implement new tools to speed up computational times. In this way, and among many others, reduced order models (ROMs) based on the proper orthogonal decomposition are one of the most popular strategies to improve efficiency and without compromising the accuracy of the solutions. In this work, we study the combination of ROMs applied to the 1D SWE with augmented Riemann solvers.

Poster 6

Neural Networks ansatz for solving quantum Long-range models

Manuel Gallego[1,2], Sebastián Roca[2], Jesús Carrete[2], David Zueco[2]

1. Department of Theoretical Physics, University of Zaragoza.
2. Instituto de Nanociencias y Nanomateriales de Aragón (INMA), CSIC.

Corresponding author: Manuel Gallego: manuel.gallego@unizar.es

In the realm of many-body systems, analytical solutions for the Schrödinger equation are often elusive. The conventional brute force method for determining eigenstates faces limitations in scalability, constrained by the exponential growth of the Hilbert space with the system size. Addressing this challenge, a recent innovative approach leveraging machine learning has emerged. This approach capitalizes on the insight that wave function coefficients, governed by symmetries within the problem, are not entirely independent. Accordingly, one tries to embed of the Hilbert space into a Neural Network's feature space which only scales polynomially with the system size. Applying this technique, we investigate the Long-range Ising model with a Transverse field. Additionally, we introduce a method for extracting phase transitions directly from the feature space of the problem.

Poster 7

Researcher identification and disambiguation across scientific production repositories

David Muñoz Jordán[1], *Alejandro Rivero Gracia*[1], *Gonzalo Ruiz Manzanares*[1]

1. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

Corresponding author: David Muñoz Jordán. email address: dmunoz-at-bifi.es

When researchers sign a publication, apply for a project, participate in a patent, etc., they must provide his personal data, say name, surname, institution where they belong, location, address, etc. Depending on the type of research activity, they may even provide more personal information such as date of birth, gender, position, ResearcherID, Scopus ID, etc. These data are usually stored in different scientific production repositories such as WOS or SCOPUS for publications, PATSTAT for patents, CORDIS for projects, etc. But it is very well known that when researchers provide their personal information, they do not always do it the same way. Over the course of the researcher's career, they may belong to different institutions, their signature may vary, they may use different personal identifiers, etc. This situation makes it really difficult to correctly identify a researcher across scientific production repositories, or even within the same one. In this poster we are going to show how natural language processing and artificial intelligence techniques can help us solve these identification problems.

Poster 8

Pitch Control Insights from Individual Velocities and Stamina Factors

Marco Fernández da Silva[1,2], Alberto Aleta[1,2], Álvaro Novillo[3] , Yamir Moreno[1,2] , Javier M. Buldú[3]

1. Department of Theoretical Physics, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Complex Systems Group, Universidad Rey Juan Carlos, Móstoles, Madrid, Spain

Corresponding author: marco.fernandez@unizar.es

Football analysis has become an increasingly prominent field, with many clubs incorporating data analysts into their staff to gain a competitive edge. However, a significant challenge arises due to the infrequency of the most pivotal event in football – the goal. That's the reason why analysts came up with alternative metrics such as pitch control, which is defined as the likelihood of a team gaining possession of the ball at any given position on the field.

In this study, we utilize player tracking data from the 2019/2020 La Liga season, capturing the position of each player at a rate of 25 frames per second to estimate pitch control. The central aspect of our approach involves considering individual velocities – something seldomly done - and adjusting a team's velocity by multiplying it with a stamina factor. This modification allows us to observe how variations in team speed impact pitch control dynamics when compared to their opponents. What we have seen is that there are certain teams less susceptible to changes in the velocity. We intend to employ graph neural networks on this dataset to devise strategies aimed at enhancing the contribution of attacking players to pitch control.

Poster 9

Predicting Non-Trivial Lipid Properties Using Neural Networks

D. Luna Cerralbo [1,2], I. Blasco Machín[3], E. Broset[3], J. Martinez[3] and P. Bruscolini[1,2]
1. Department of Theoretical Physics, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Department of RNA Design, Certest Pharma.

Corresponding author: David Luna Cerralbo: luna.cerralbo.david@gmail.com

The direct inference of non-trivial lipid properties, such as the logP or the Molar Refractivity, from their molecular structure is a challenge, and existing and widely recognized tools, such as RDKit and OpenBabel, often fall short in accurately predicting these intricate characteristics. However, these tools successfully calculate more straightforward lipid properties, like the fraction of heteroatoms or the number of rings. We compute these "trivial" properties as intermediate step and use them as input for a neural network.

Our methodology involves harnessing the predictive power of a neural network, fine-tuned on these intermediate properties, to accurately forecast a spectrum of non-trivial lipid characteristics. The study not only showcases the effectiveness of this approach in circumventing the limitations of existing tools, but also provides insights into the successful prediction of complex lipid properties, shedding some light on the intricate relationship between lipid structure and function.

Poster 10

Inferring effective couplings with Restricted Boltzmann Machines

A. d. J. Navas Gómez[1], A. Decelle[1,2], C. Furtlehner[2], B. Seoane[2]

1. Departamento de Física Teórica, Universidad Complutense de Madrid, 28040 Madrid, Spain.
2. Université Paris-Saclay, CNRS, INRIA Tau team, LISN, 91190, Gif-sur-Yvette, France

Presenting author: Alfonso de Jesús Navas Gómez: alfonn01@ucm.es

Generative models offer a direct way to model complex data. Among them, energy-based models provide us with a neural network model that aims to accurately reproduce all statistical correlations observed in the data at the level of the Boltzmann weight of the model. However, one challenge is to understand the physical interpretation of such models. In this study, we propose a simple solution by implementing a direct mapping between the energy function of the Restricted Boltzmann Machine and an effective Ising spin Hamiltonian. This mapping includes interactions of all possible orders, going beyond the conventional pairwise interactions typically considered in the inverse Ising (or Boltzmann Machine) approach, and allowing the description of complex datasets. To validate our method, we performed several controlled inverse numerical experiments where we trained the RBMs using equilibrium samples of predefined models containing local external fields, 2-body, and 3-body interactions in various low-dimensional topologies. The results demonstrate the effectiveness of our proposed approach in learning the correct interaction network and pave the way for its application in modeling interesting binary datasets. We also evaluate the quality of the inferred model based on different training methods.

Poster 11

Integrating Social Determinants and Machine Learning to Understand COVID-19 Mortality in Catalonia, Spain

Charles Novaes de Santana[1], Juan Fernández-Gracia[1], Jorge P. Rodríguez[1], **Alberto Aleta**[2,3], Álvaro Padilla-Pozo[4,5], Víctor M. Eguíluz[6], Francesc Belvis[7,8], Joan Benach[7,8,9] and the COVID-SHINE group.

1. Instituto de Física Interdisciplinar Y Sistemas Complejos IFISC (CSIC-UIB), Palma de Mallorca.
2. Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza.
3. Department of Theoretical Physics, University of Zaragoza, 50018, Zaragoza, Spain.
4. Department of Sociology, Cornell University, Ithaca, NY, U.S.A.
5. Cornell Population Center, Cornell University, Ithaca, NY, U.S.A.
6. Basque Centre for Climate Change (BC3), Leioa, Spain
7. Research Group on Health Inequalities, Environment, and Employment Conditions (GREDS-EMCONET), Department of Political and Social Sciences, Univ. Pompeu Fabra.
8. Johns Hopkins University-Universitat Pompeu Fabra Public Policy Center (JHU-UPF PPC).
9. Ecological Humanities Research Group (GHECO), Universidad Autónoma de Madrid.

Corresponding author: Alberto Aleta (aaleta@unizar.es)

This study investigates the association between socioeconomic conditions and COVID-19 mortality in hospitalized patients in Catalonia, Spain. Our dataset includes 42,033 COVID-19 hospitalizations from June 21, 2020, to December 6, 2020. A XGBoost classification model is trained using an 80-20 dataset split. Model performance is evaluated using the AUC and SHAP values identify the most influential features related to COVID-19 mortality. Simulation experiments assess the impact of changes in social conditions on mortality rates.

The XGBoost classification model demonstrates strong discriminatory power, achieving an AUC value of 0.8567, accurately predicting COVID-19 mortality in hospitalized patients in Catalonia. Age, comorbidities risk index, income, ICU admission, and gender are significant predictors of mortality, highlighting the multifaceted nature of disease outcomes. Notably, lower income levels are associated with an increased risk of death, emphasizing the impact of socioeconomic disparities on disease severity. Simulation experiments suggest that worsening social conditions alone might not substantially alter mortality rates among hospitalized patients. Conversely, improving social factors, particularly income disparities, show potential benefits in reducing COVID-19 mortality.

Our study provides valuable insights into the complex interplay of socioeconomic factors and COVID-19 mortality in Catalonia. The machine learning analysis and simulation experiment underscore the importance of considering social determinants in formulating effective public health policies and interventions.

Poster 12

"Knowledge Search Engine": linking the gap between Academy and Industry

Juan Luis Durán Batalla[1], Óscar Fraca[1], Alejandro Rivero[1,3], Alfonso Tarancón[1,2,3]

1. Kampal Data Solutions.
2. Department of Theoretical Physics, University of Zaragoza.
3. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

*Corresponding author: Juan Luis Durán Batalla: jduran-*at*-kampal.com*

Addressing complex challenges necessitates specialized expertise. The industry often harbors a diverse array of problems that require the attention of researchers. Yet, bridging the gap between academia and industry to connect with experts has historically posed significant challenges—until now. Earlier this year, Kampal Data Solutions introduced a groundbreaking "Knowledge Search Engine"[4] designed to facilitate this connection. This innovative search engine extracts keywords from academic papers and projects, serving as a gateway to identifying pertinent researchers. Central to this algorithm is a powerful Large Language Model (LLM) that, together with various NLP tools, effectively extracts keywords from textual content. This integration allows for fast multilingual retrieval within the search engine itself. The resulting application is presented via a user-friendly interface, aimed at fostering stronger collaborations between universities and industry.

[4] https://buscadordeconocimiento.unizar.es/buscador-conocimiento-search.html

Poster 13

Hydrogen Bonding Patterns and Cooperativity in Polyproline II Helical Bundles

Rubén López-Sánchez[1], Douglas V. Laurents[1], Miguel Mompeán[1]

1. Instituto de Química Física "Blas Cabrera" - Consejo Superior de Investigaciones Científicas (IQF-CSIC).

Corresponding author: Rubén López-Sánchez: rlopez@iqf.csic.es

Hydrogen bond cooperativity (HBC) plays an important role in the stability of protein assemblies built by α-helices and β-sheets [1], which are the most common protein secondary structure building blocks. However, whether HBC exists in other types of protein secondary structures such as polyproline II (PPII) helices remains unexplored. Understanding the bases of PPII helical stability is important since their assemblies are increasingly found in natural systems and find numerous applications in engineered proteins and chemical biology. Here, using a combination of computational chemistry tools and molecular modeling corroborated by experimental observables [2], we establish that HBC stabilizes intermolecular PPII helices as seen in other types of protein assemblies such as amyloid fibrils. In addition to cooperative interactions in canonical CO···HN H-bonds, we show that analogous interactions in non-canonical CO···$H_\alpha C_\alpha$ H-bonds are of particular relevance in Gly-rich PPII bundles, thus compensating for the inability of Gly residues to create hydrophobic cores. Our results provide a mechanistic explanation for the assembly of these bundles.

Acknowledgements:
This work has been funded by grants PID2019-109306RB-I00, PRE2020-094019 and PID2022-137806OB-I00 from MCIN/AEI/10.13039/501100011033/FEDER, UE (D.V.L.).

References:

1. Mompeán, M.; Nogales, A.; Ezquerra, T. A.; Laurents, D. V. Complex System Assembly Underlies a Two-Tiered Model of Highly Delocalized Electrons. J. Phys. Chem. Lett. 2016, 7 (10), 1859–1864.
2. Treviño, M. Á.; Pantoja-Uceda, D.; Menéndez, M.; Gomez, M. V.; Mompeán, M.; Laurents, D. V. The Singular NMR Fingerprint of a Polyproline II Helical Bundle. J. Am. Chem. Soc. 2018, 140 (49), 16988–17000.

Poster 14

Fluorescence Spectroscopy Combined with Nanoparticles and Machine Learning Analysis for the Diagnosis of Tumoral Pathologies

Oscar Sanchez-Gracia[1], Violeta Morcuende-Ventura[2], Natalia Abian-Franco[3], Sonia Hermoso[1,4,5,6], Alfonso Ortega[7], Teresa Sierra[2], Olga Abian[1,4,5,6,8], Adrián Velázquez-Campoy[1,4,5,6,8]

1. Universidad de Zaragoza, Spain
2. Instituto de Nanociencia y Materiales de Aragón (INMA), Química Orgánica, Facultad de Ciencias, CSIC-Universidad de Zaragoza, Spain
3. Hospital Infanta Sofia, Tudela, Spain
4. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Spain.
5. Aragón Health Research Institute (IIS-Aragón), Zaragoza, Spain.
6. CIBERehd, Spain
7. ViVoLab, Aragón Institute for Engineering Research (I3A), University of Zaragoza
8. Department of Biochemistry and Molecular and Cell Biology, University of Zaragoza, Spain.

Corresponding author: Oscar Sanchez-Gracia: oscar.sanchez.gracia@gmail.com

Introduction: The differentiation of malignant and benign biological samples with tumoral pathologies remains a critical challenge in clinical diagnostics. Our study explores the potential of fluorescence spectroscopy enhanced with dendrimer-based nanoparticles, combined with Machine Learning (ML) algorithms, to advance the diagnostic accuracy for ovarian, pancreatic, and colorectal cancer.

Materials and Methods: We analyzed serum samples from patients with ovarian, pancreatic, and colon cancers using three types of dendrimer-based nanoparticles to enhance fluorescence spectroscopy signals. Advanced ML models, including AdaBoost, Random Forest, CatBoost, Decision Tree, Gaussian NB, and KNeighbors Classifier, were employed to analyze the differential fluorescence spectra of patients versus control groups.

Results: The ML models demonstrated promising diagnostic capabilities, with AUCs consistently above 0.8 for specific nanoparticles and pathologies. The differential analysis provided by the fluorescence spectra, enhanced by the nanoparticles, showed distinct patterns that could effectively distinguish between malignant and benign conditions.

Conclusions: The integration of dendrimer-based nanoparticle-enhanced fluorescence spectroscopy with sophisticated ML algorithms offers a novel and promising approach to the diagnosis of different types of cancers. The use of easily accessible serum samples and the high accuracy of the models suggest a significant potential for this methodology in clinical settings. Further validation with larger sample sizes and diverse patient populations is envisaged to establish the robustness of this approach.

Poster 15

Integration of Machine Learning with Thermal Liquid Biopsy (TLB) using serum samples and patient-specific variables analysis for presurgical diagnosis of malignant ovarian cysts

Francisco Javier Falcó[1], Natalia Abian-Franco[2], Sonia Hermoso[1,3,4], Nicolas Fraunhoffer[5], Adrián Velázquez-Campoy[1,3,4,6], Olga Abian[1,3,4,6]

1. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Spain.
2. Hospital Infanta Sofía, Tudela, Spain
3. Aragón Health Research Institute (IIS-Aragón), Zaragoza, Spain.
4. CIBERehd, Spain
5. Centre de Recherche en Cancérologie de Marseille (Crcm), France.
6. Department of Biochemistry and Molecular and Cell Biology, University of Zaragoza, Spain.

Corresponding author: Francisco Javier Falcó Martí: ffalco@unizar.es

Introduction: Screening for malignancy of ovarian cysts requires new tools to avoid unnecessary surgeries in non-malignant cysts. The integrated use of Thermal Liquid Biopsy (TLB) of serum samples is proposed, together with demographic and biochemical data, using Machine Learning (ML) algorithms, with the aim of restricting surgery exclusively to necessary cases of malignant ovarian cysts.

Materials and Methods: Serum samples were analyzed by TLB and a temperature pair-based analysis of thermograms (k–Top Scoring Pairs classifications), with 100-fold repetitions, k-fold cross-validation and variables regularization methods, has been used for the development of a malignancy prediction model. Statistical analyses have elucidated the relationship between different variables and ovarian cyst malignancy.

Results: The models generated have returned Areas Under the Curve above 0.75 and sensibility values around 80%. Menopausal status and smoking, among other variables, have shown a statistically significant relationship with malignancy in ovarian cysts.

Conclusions: The combined use of TLB from serum samples with clinical variables, applying ML algorithms, has proven to be a useful tool for the determination of ovarian cyst malignancy. Nevertheless, further research using larger sample populations could improve the diagnostic capacity of this promising technique for the presurgical diagnosis of patients with ovarian cysts.

Poster 16

Towards machine learning based diagnosis of Parkinson's disease from two-color fluorescence detection of amyloid aggregates in human biofluids

David Polanco[1,*], Alejandra Carrancho[1], María Martínez-Monge[1], Nunilo Cremades[1,2]

[1] Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain

[2] Department of Biochemistry and Molecular and Cell Biology, University of Zaragoza, 50009 Zaragoza, Spain

Corresponding author: David Polanco dpolanco@unizar.es

Diagnosis of Parkinson's disease (PD) traditionally relies on medical observations of clinical signs such as motor symptoms and brain imaging. Nonetheless, these approaches can only be applied in patients with advanced stages of the disease. Our framework aims for the development of an early diagnosis by detecting pathological amyloid aggregates of alpha-synuclein (αSyn), the main hallmark of PD, in human biofluids with single-molecule fluorescence (SMF) methods employing fluorescently labelled probes with high affinity to αSyn aggregates. These aggregates have multiple bindings sites for the chosen probes, and therefore the particle fluorescence intensity is typically higher than the background of unbound probe, although the distribution of intensities of the aggregates and the background can overlap, making it difficult to discriminate well the entire distribution of aggregate sizes from the background. Here, we show a preliminary attempt to resolve this issue by using a machine learning (ML)-based approach to identify the optimum fluorescence intensity thresholds for the discrimination between positive samples containing αSyn aggregates (intensity bursts + background) and a negative one (background only). We also want to expand this ML methodology for more sample descriptors to be finally applied for a high-throughput diagnosis of real samples.

Keywords: machine learning, single molecule fluorescence, Parkinson's disease, alpha synuclein, deep learning diagnostics.

Poster 17

Glycine-Rich Polyproline II Helical Bundle Domains Characterized for Improving AI Protein Structure Prediction

Cristian Segura Rodríguez[1], Douglas V. Laurents[1]

1. Instituto de Química Física "Blas Cabrera" CSIC Serrano 119, 28006 Madrid

Corresponding author: Cristian Segura Rodríguez: cristian.segrod.gmail.com

Machine learning methods have succeeded brilliantly in predicting the conformation of most well-folded proteins [1]. However, their success is patchy with glycine-rich polyproline II helical bundle domains. These folded domains can be composed of >50% glycine residues which are generally predicted to be disordered but in fact form robust, ordered structures with distinct characteristics [2]. To better understand these protein domains, we have compared the several glycine-rich polyproline II helical bundle domains from several natural proteins with different functions and diverse evolutionary origins. We find that these domains: 1) are well-packed and have few cavities, 2) preferentially aligned their helices in an anti-parallel configuration which may favor macro-dipole interactions, 3) are tolerant of different connecting loops and flanking sequences and 4) are stabilized by 3D networks of N-H···O=C and Cα-H···O=C hydrogen bonds [3]. These findings could help improve machine learning methods and guide the design of novel proteins incorporating glycine-rich polyproline II helical domains.

Acknowledgement: Este estudio es parte del proyecto PID2022-137806OB-I00, financiado por MCIN/AEI/10.13039/501100011033/FEDER, UE.

[1] Jumper J *et al.* (2021) Nature **596**: 583-589.
[2] Laurents DV (2022) Front. Mol. Biosci. **9**: 906437.
[3] Segura-Rodríguez C & Laurents DV (2023) bioRvix:
https://doi.org/10.1101/2023.11.25.568672

Poster 18

A statistical-physics model for codon usage

D. Luna Cerralbo [1,2], I. Blasco Machín[3], E. Broset[3], J. Martinez[3], **P. Bruscolini[1,2]**
1. Department of Theoretical Physics, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.
3. Department of RNA Design, Certest Pharma.

Corresponding author: Pierpaolo Bruscolini: pier-at-unizar.es

The degeneracy of the codon alphabet allows different codons to translate to the same amino acid, and different species show different patterns of codon usage. Using the wrong pattern reduces protein expression; however, using the most frequent codon for each amino acid does not necessarily maximize protein expression, which is the goal of codon "optimization".

Due to the complex (and mostly unknown) factors connecting codon sequence to protein *in vivo* expression, we choose a "safe" approach to optimization, that involves adjusting the codon sequence to mimic the “natural” choice a given species would make, to produce the target protein.
To this end, we propose a simple statistical-physics model, where the probability of any codon sequence is related to the “interactions” between neighboring codons.

We infer the model's parameters by maximizing the overall probability of a large databases of human proteins. Then, we test the method by redesign the codon sequence of some proteins, and we perform experiments on Luciferase, as a handy reporter protein.

Poster 19

Digital language and communication training for EU scientists

Pérez-Llantada, C.[1,2], Carciu, O.[1,2], Villares, R.[1,2]

1. Department of English and German Studies, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

Corresponding author: C. Pérez-Llantada, llantada@unizar.es

Our main aim in this presentation is to introduce DILAN, an EU-funded Erasmus+ project whose main objective is to improve the ability of STEM scientists to communicate their science to diverse audiences. Communicating science beyond expert audiences has been recognised as a key societal priority in the EU. However, there is a lack of resources and training specifically designed to enable scientists to support this key priority. We explain how DILAN will contribute to supporting this priority by creating digital professional development resources to help STEM scientists (especially women) better communicate their science to society. We first describe the ethnomethodological study we conducted to collect data from EU STEM scientists on digital science communication practices and to create an online inventory of good practices and video testimonials from EU women scientists on how they communicate science beyond expert audiences using digital tools and media. We will also explain how we will produce teaching materials and develop a digital (online) training course and a MOOC to help EU STEM scientists communicate more effectively to public audiences in different languages, to enhance their intercultural sensitivity.

Poster 20

scDVar: a computational tool for modeling inter-individual variation of cell-to-cell transcriptional noise in single-cell RNA-seq.

Ignacio Marchante[1,2], Joaquín Sanz[1,2]

1. Department of Theoretical Physics, University of Zaragoza.
2. Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza.

Since its birth in 2009, single cell RNA sequencing technologies (scRNA-seq) have contributed to the development of multiple areas of Biology, enabling analysis of complex tissues in unprecedented detail. Among the many exciting new avenues paved by this new technology, inspecting scRNA-seq transcriptomes unlocks the characterization of cell-to-cell transcriptional heterogeneity, and how this varies across genes, within a single sample, or, for the same gene, across tissues and individuals.

However, studying variation in scRNA-seq requires careful data handling and interpretation, as cell-to-cell variation shows a strong dependence to mean gene expression -heteroskedasticity-, but also to sequencing depth. Furthermore, dispersion of cell-to-cell variances around their expected values, also depends on mean expression. As a result of these difficulties, most applications of variance modeling in scRNA-seq data focus on the comparison of variance across genes, within sample (for the sake of identification of simple-specific highly variable genes), but fail short when applied to the comparatively harder task of identifying differentially variable genes (DVGs) across subjects.

In this work, we present scDVar, a computational pipeline that enables cell-to-cell gene expression variance modelling and differential variance analyses in scRNA-seq data. Our method relies on generalized additive models for location, scale and shape (GAMLSS) to capture non-trivial dependences of variance distributions to mean expression, yielding a proper management of the main sources of confounding; and implements a bootstrap procedure through which differences in confidence in variance estimates across individuals sampled heterogeneously are translated into precision weights.

Applying our method to a recently published dataset on the effects of African vs. European genetic ancestry on immune responses to infections [1] we show how in-vitro infection effects on cell-to-cell expression variance is highly asymmetric across cell types; and identify genes whose levels of transcriptional coherence depend on ancestry without a significant shift in their mean expression profiles.

References

[1] Randolph, HE., et al. Science 374.6571 (2021)

Poster 21

Comparison between classical linear regression and neural networks for the critical temperature of superconductors

Alcalde, Martín[1]

1. Department of Statistical Methods, University of Zaragoza.

Corresponding author: malcalde@unizar.es

In Physics, superconductivity is the property of some materials that enables them to conduct electric current without opposing resistance after lowering its temperature to a certain temperature, which is called critical temperature. Nowadays, no mathematical formulas for the prediction of the critical temperature are known, so we use two statistical models for its prediction from several physical properties of materials. Our aim is to show that neural networks can be more useful for predicting the critical temperature than classical linear regression.

Figure References

Book cover: Maitane Gutierrez.

Back cover image. OpenAI. 2023. "Background image depicting a futuristic and highly technological scenario, inspired by artificial intelligence" DALL·E. Accessed [30 december 2023].

AI Methods and Applications Cover Image. OpenAI. 2023. " AI Methods and Applications" DALL·E. Accessed [30 december 2023].

AI and Statistical Modelling Cover Image. OpenAI. 2023. “AI and Statistical Modelling. high-tech background with glowing digital patterns, symbol” DALL·E. Accessed [2023-12-30].

AI Applications in Physics Cover Image. OpenAI. 2023. “AI applications in physics, without any text”. DALL·E. Accessed [2023-12-30].

AI Applications in Biomedicine Cover Image. OpenAI. 2023. “A sophisticated and futuristic image representing 'AI Applications in Biomedicine'.” DALL·E. Accessed [2023-12-30].


Instituto Universitario de Investigación
Biocomputación y Física
de Sistemas Complejos
Universidad Zaragoza
Universidad
Zaragoza
1542
kampal
certest
DELL
Technologies
intel