

[SORS/WomenInBSC: Exploring Neurodegenerative Diseases in the Era of Machine Learning](#)

Abstract

The presentation integrates three studies addressing different aspects of neurological diseases using machine learning (ML). The first study implements network analysis, protein language models (pLMs), and deep learning to address key aspects of Amyotrophic Lateral Sclerosis (ALS), recognizing the lack of reliable biomarkers and effective treatments. We employed a network-based classification method using graph convolutional neural networks (GCNNs) to integrate gene expression profiles with protein-protein interaction (PPI) networks. We analyzed two ALS microarray datasets alongside three PPI networks and compared the performance of the GCNN approach with network-free methods, such as logistic regression, random forest, and XGBoost. The GCNN and logistic regression models demonstrated good performance as classifiers for ALS patients, whereas the other tested methods did not. By applying Graph Layer-wise Relevance Propagation (GLRP) to the GCNN, we identified a highly stable set of relevant genes across all networks, which were enriched in the same biological pathways.

The second study addresses the protein aggregation phenomenon as a pathological hallmark of several neurodegenerative diseases. It developed a deep learning-based predictor for aggregation-prone regions (APRs) using ProtT5 embeddings on curated amyloid databases (CPAD and AmyPro), as well as homology-based predicted APRs. This novel approach, incorporating class balancing and cross-validation, demonstrated comparable or superior performance to existing methods that rely solely on primary sequence information, highlighting the utility of pLMs in APR prediction.

Finally, we explored the impact of disease-related mutations in heterogeneous nuclear ribonucleoproteins (hnRNPs), which are implicated in several neurological diseases (including ALS, Parkinson's disease, and Alzheimer's disease) and are also known for their role in liquid-liquid phase separation (LLPS) and aggregation. Using the ProtT5 pLM, we generated embeddings for wild-type and mutated hnRNP sequences from ClinVar, as well as for all possible single amino acid substitutions, to assess the ability of pLMs to capture the effects of these mutations and, potentially identify novel disease-causing mutations in other LLPS proteins.

These combined computational approaches provide valuable insights into ALS pathogenesis, from gene expression and network-level perturbations, the impact of mutations and protein aggregation, offering potential avenues for biomarker discovery and therapeutic development.



Short Bio

Cristina Marino-Buslje is a Principal Investigator at CONICET and a Professor at the Faculty of Bioengineering at the Buenos Aires Institute of Technology (ITBA). She has led the Structural Bioinformatics Unit at the Fundación Instituto Leloir in Buenos Aires, Argentina, since 2010. Dr. Marino-Buslje earned her Ph.D. in Biological Sciences from the Autonomous University of Barcelona (UAB) in 1996, following a Master's in Biotechnology and an undergraduate degree in Biology from the same institution. She furthered her research with a postdoctoral Marie Curie fellowship at the University of Cambridge, UK and expanded her research experience as a visiting professor at the Tata Institute of Fundamental Research in Bangalore, India. She was a founding member at 2010 and former President of the Argentine Association of Bioinformatics (A2B2C). Dr. Marino-Buslje's research investigates the structure, function, evolution, and interactions of proteins to elucidate biological processes, particularly those implicated in disease. Her research interests encompass intrinsically disordered proteins and phase-separating proteins. Dr. Marino-Buslje has a strong track record in developing bioinformatics tools for analyzing biological data. Recently, her research has integrated machine learning methodologies, including the application of Convolutional Neural Networks (CNNs) neurodegenerative diseases analysis and the use of embeddings for studying RNA-binding proteins and aggregation-prone regions in amyloid-forming proteins among other research.

Speakers

Speaker: Cristina Marino-Buslje. Principal Investigator at CONICET and Professor at the Faculty of Bioengineering at the Buenos Aires Institute of Technology (ITBA)

Host: Gonzalo Parra. Established Researcher, Computational Biology Life Sciences Group, Life Sciences Department, BSC

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 2 Abr 2025 - 15:05): <https://www.bsc.es/es/research-and-development/research-seminars/sorswomeninbsc-exploring-neurodegenerative-diseases-the-era-machine-learning>