Titles and abstracts:

Speaker: Vincent Calvez, CNRS- Université Claude Bernard
Title: Self-generated gradients in biology: A case study.
Abstract: I will present a simple biological experiment of amoeboid cells confined in hypoxic conditions, with a simple, yet remarkable outcome: a ring of cells traveling over days at constant speed with a constant shape. Then, I will present a simple PDE model that recapitulates the biological knowledge and agrees with the experimental data.

Speaker: Ezequiel Treister, Pontificia Universidad Católica de Chile
Title: Data-Driven Science in Astrophysics in the Context of the Center for Astrophysics and Related Technologies
Abstract: In this talk I will introduce the Center for Astrophysics and Related Technologies (CATA), which was recently recommended for renewal by the National Agency for Research and Development (ANID). In particular, I will focus on the initiatives proposed for the next 5 years related to data-driven science in astrophysics. Among them, I will describe galaxy evolution simulations carried out in Chile, efforts to store and analyze the vast amounts of data generated by the international observatories installed in Chile, and work using state-of-the-art Machine Learning algorithms to analyze the imaging data that will be produced by the Legacy Survey of Space and Time (LSST) to be carried out at the Vera Rubin Observatory starting in 2024.

Speaker: Vittorio Perduca, Laboratoire MAP5 (UMR CNRS 8145), Université de Paris
Title: Target learning in the time of Covid-19: a French case study
Abstract: The primary goal of this project is to develop a disease severity predictor for patients hospitalized for Covid-19. In particular, we leverage on demographic, biological and clinical variables, to predict disease remission or admission to intensive care. For this, rather than arbitrarily choosing a machine learning method, we rely on the super learner, an ensemble learning method that allows to rigorously evaluate and compare the performance of basic learning algorithms in a user-provided library. The secondary goal is to define and estimate the importance of explanatory variables to better understand their real relevance in the evolution of the disease. For this, we use a measure of importance introduced in the literature of targeted learning. We apply the super learner and our measure of the importance of variables to data collected during the first wave at Hôpital Bicêtre in the Spring 2020.

Speaker: Jocelyn Dunstan, Universidad de Chile.
Title: Natural language processing to support clinical decisions
Abstract: The clinical text analysis has particular challenges due to the extensive use of non-standardized abbreviations, the variability of the clinical language across medical specialties and health professionals, and its restricted availability for privacy reasons, to mention some. The challenge is even steeper when working on languages other than English because of the limited availability of tools and training corpora. Natural Language Processing (NLP) is a branch of Artificial Intelligence focused on designing methods and algorithms that take as input or produce as output data in the form of human language. This talk will present the application of NLP to two relevant tasks in Chilean public health. The first one is an automatic classification of patients waiting for a specialty consultation. The second is identifying key information mentioned in a referral, such as diseases, body parts, or medications. An appropriate machine interpretation of medical narratives in Chile can

unleash information and knowledge in large volumes of medical narratives, expanding clinical management and facilitating the secondary use of information for research purposes.

Speaker: Emille Ishida, Université Clermont-Auvergne
Title: Fink broker, an optimize recommendation system for transient follow-up
Abstract: Next generation experiments such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) will provide an unprecedented volume of time-domain data opening a new era of big data in astronomy. To fully harness the power of these surveys, we require analysis methods capable of dealing with large data volumes that can identify promising transients within minutes for follow-up coordination. In this talk I will present Fink, a broker developed to face these challenges. Fink is based on high-end technology and designed for fast and efficient analysis of big data streams. It has recently been chosen as one of the official LSST brokers and will receive the full data stream. The project will be hosted at CC-IN2P3 and provide unique opportunities for users with LSST data rights. I will highlight the state-of-the-art machine learning techniques used to generate early classification scores for a variety of time-domain phenomena including supernovae and microlensing events. Such methods include Deep Learning advances and Active Learning approaches to coherently incorporate available information, delivering increasingly more accurate added values throughout the duration of the survey. Such results will certainly be made significantly more informative when cross-matched LSST catalogs and provide an ideal environment for identification of interesting LSST transients and subsequent follow-up.

Speaker: Rodrigo Abarca Del Río (University of Concepción)
Title: Environmental and geophysical analysis utilizing advanced data processing techniques in remotely sensed data: Research conducted at the Universidad de Concepción
Abstract: Sensor technology advancements, an ever-increasing number of satellite missions, and available computing power have all aided the rise of remote sensing applications. Processing sophisticated, multi-scale, multi-dimensional data from drones, satellites, environmental sensors, or even climate model simulations, on the other hand, presents numerous challenges, particularly for advanced image processing methods that convert massive amounts of spectral data into decision-support information. Combining machine learning/artificial intelligence with powerful computational processing (local or in the cloud via big data) allows for the automatic extraction, analysis, and interpretation of meaningful and relevant information.
The Remote Sensing and Data Analysis Group is part of the University of Concepción's Faculty of Physics and Mathematics and is affiliated with undergrad and postgraduate programs (in Geophysics, Environmental Sciences, Physics, Energy, Geology). The laboratory specializes in earth observation. Its research and application focus is on environmental phenomena centered on energy, water, and pollution in natural and urban settings. Applications include energy estimation using satellite and radar sensors, observation of the various components of the water cycle (snow, glaciers, water) and reservoirs (rivers, lakes, aquifers, etc.), and the development of new statistical and mathematical methodologies (for ex; inversion methods to extract geophysical parameters or geometric features from satellite images for water contamination issues). We will take a quick tour of several of the group's applications during the presentation.

Speaker: Clémentine Prieur, Université Grenoble-Alpes

Title: Sensitivity analysis and parameter estimation

Abstract: In most application fields, analysts support their investigations through the use of quantitative models. Over the years, the computation of sensitivity (or importance) measures has become an integrating part of the analysis, allowing analysts to obtain insights on key drivers of model behavior either on a local or on a more global scale. Therefore the efforts can be focused on the estimation of the parameters revealed as key parameters by sensitivity analysis. The aim of this short talk is to point out some of the challenges in this area.

Speaker: Carlos Sing-Long, Pontificia Universidad Católica de Chile
Title: Random artifacts and its potential effects in medical image reconstruction
Abstract: The application of learning techniques for medical image classification has attracted the interest of the community in the past years. Although these techniques have the potential to impact medical practice, recent studies have shown that some of them, e.g., neural networks, can leverage spurious image artifacts to achieve low validation error on a given dataset. This precludes their widespread adoption. One potential source of such artifacts are image reconstruction techniques. The artifacts in the reconstructed images can arise from reconstruction errors and from changes in the statistics of the measurement noise. Although for linear methods both types of artifacts can be characterized, for non-linear methods the problem is challenging. In this short talk we discuss our recent efforts aiming to characterize the probability distribution of the noise on the reconstructed image in 4D flow magnetic resonance imaging, focusing on the presence of spatial correlations. We consider Compressed Sensing as the reconstruction technique due to its widespread use, its rich mathematical foundations, and its theoretical guarantees.

Speaker: Jerome Bolte, ANITI Toulouse
Title: Nonsmooth automatic differentiation: the key to the learning phase of Neural Networks
Abstract: Nonsmooth automatic differentiation as implemented through modern libraries as TensorFlow or PyTorch, is an essential ingredient in the training of deep neural networks and in the AI revolution. I will give an insight into this question and I will also evoke the recent mathematical model we provided for this widespread practice.

Speaker: Francisco Förster, Universidad de Chile.
Title: The Universe in a Stream: Challenges and Progress of the ALeRCE Broker
Abstract: A new generation of large aperture and large field of view telescopes is allowing us to explore large volumes of the Universe in an unprecedented fashion. This has led to the discovery of new populations of astrophysical events or new phases of evolution of known objects. In order to take advantage of these new telescopes a new time domain ecosystem is developing. Among the tools required for this new ecosystem are fast machine learning aided discovery and classification algorithms, interoperable tools that allow for an effective communication between brokers and follow-up telescopes for rapid reaction, and new models and tools to extract the most physical knowledge from these observations. In this talk I will review the challenges and progress of building one of these systems: the Automatic Learning for the Rapid Classification of Events (ALeRCE) astronomical alert broker. ALeRCE (http://alerce.science/) is a new alert annotation and classification system led by an interdisciplinary and interinstitutional group of scientists from Chile and the US. ALeRCE is focused around different scientific cases: stellar explosions, variable stars, supermassive black holes, and asteroids. I will discuss some of the challenges associated with the problem of alert classification, including the ingestion, annotation, database management, training set building, distributed processing, machine learning classification

and visualization, or the challenges of working in large interdisciplinary teams. I will show some results based on the real-time ingestion and classification using the Zwicky Transient Facility (ZTF) alert stream as input, as well as some of the tools available. In about one year of operations we have ingested more than 120 million alerts, classified about 30 million objects based on their images, classified about one million objects based on their light curves, and reported more than 4000 supernova candidates.