ASTROINFORMATICS 2023 PROGRAM

October 1 Welcome cocktail and pre-registration 18:30 - 20:00

October 2 MORNING SESSION

	9.15-9.30	*Welcome address and communications from the LOC*			
Keynote	9.30-10.30	Yann LeCun	NYC & Facebook	AI-powered future of imaging science	tba
Invited	10.30-11	Matthew Graham	Caltech (USA)	Developments in Fast Machine Learning for Science	For over two decades, we've heard about the challenges that the forever imminent data tsunami is going to bring to astronomy and yet we still spend most of our computing lives firmly in the land of CPUs with occasional forays into the GPU realm. There are, however, now genuine needs across the entire science spectrum for high throughput low latency inferencing driven by real-time responses that our current systems cannot deliver. Novel options in both hardware components and algorithmic methodologies are required to deliver production-ready solutions that can cope. In this talk, I will review current developments in the field of fast machine learning and, in particular, those that are relevant to multimessenger astronomy which is at the forefront of this work in our domain.
Coffee break	*11.00-11.30*				
Invited	11.30-12.00	Pavlos Protopapas	Harvard (USA)	Physics Informed Neural Networks (PINNS); Solving Differential Equations Using Neural Networks.	In this talk, I will review how we can use neural networks to solve differential equations for initial conditions and boundary value problems. There are many exciting directions in this line of work, but two main criticisms are computational complexity compared to traditional methods and the lack of error bounds. I will present transfer learning methods that can be used to speed up computation and new work on estimating the error bounds of the neural network approach in solving (partial) differential equations.
contributed	12.00 - 12.20	Michele delli Veneri	INFN- Napoli	Deep Focus and ALMASim – A meta-learner for the resolution of inverse problems and its companion simulation package	Current and forthcoming astronomical observatories are rapidly increasing the quantity, velocity and complexity of their data products transitioning in the Big Data regime. Extracting knowledge from scientific data produced by such instruments involves the resolution of ill-posed inverse problems traditionally solved with algorithms which cannot cope anymore with the rising data complexity. We present Deep Focus (DF), a meta-learner for the resolution of inverse problems in astrophysical imaging. The method can construct and optimize Deep Learning (DL) architectures (Res-Nets, U-Nets, VGG-Nets, Dense-Nets, Autoencoders) from a large pool of model hyperparameters. Leveraging multi-GPU environments, and a Bayesian hyperparameter optimization strategy, DF can search for the most performing architecture for a given problem.We also present ALMASim, a simulation package written in Python for the Atacama Large Millimeter/Submillimeter Array (ALMA), capable of rapidly generating thousands of mock ALMA observations of both point-like and extended sources. The package can produce, given a set of real observations, a dataset of statistically similar mock data which can be used to train and test DL algorithms. This is performed by sampling, for similar observing conditions, from a distribution of closely related source morphologies and noise profiles. We tested the performances of DF in deconvolving 3D ALMA mock data cubes generated by ALMASim, containing high-redshift point-like sources (calibrators) surrounded by fainter serendipitous sources. We obtained better reconstruction capabilities than those obtained with other "standard: DL architectures and achieved two order of magnitude improvements in speed and reconstruction quality with respect to the tCLEAN algorithm.

contributed	12.20 - 12.40	Ekaterina Govorkova	Massachussetts Institute of Technology	GWAK: Gravitational-Wave Anomalous Knowledge with Recurrent Autoencoders	Deep Learning assisted Anomaly detection is quickly becoming a powerful tool allowing for the rapid identification of new phenomena. We present a method of anomaly detection techniques based on deep recurrent autoencoders to the problem of detecting gravitational-wave (GW) transients in laser interferometers. This class of algorithm is trained via a semi-supervised strategy, i.e. with a weak distinction between classes at training time. While the semi-supervised nature of the problem comes with a cost in terms of accuracy as compared to supervised techniques, there is a qualitative advantage in generalizing experimental sensitivity beyond pre-computed signal templates. We construct a low-dimensional embedded space we refer to as GWAK (Gravitational-Wave Anomalous Knowledge) which captures the physical signatures of distinct signals on each axis of the space. By introducing alternative signal priors that capture the salient features of gravitational-wave signatures, we allow for the recovery of sensitivity even when an unmodelled anomaly is encountered. We show that regions of the embedded space can identify compact binary coalescences, sine Gaissian-like short duration signals and detector artifacts (glitches), and also search a variety of hypothesized astrophysical sources that may emit short duration signals in the GW frequency band including core-collapse supernovae and other unmodelled transient sources. Proved to be efficient, we incorporate the GWAK search pipeline as a part of the ML4GW software stack. We show how the ML4GW stack is quickly becoming an effective toolkit for the fast and effective deployment of Machine Learning based gravitational algorithms.
Panel discussion	12.40- 13.30	Chair: G.S: Djorgovski		The future of Astroinformatics	
Lunch break	*13.30-14.30*				
	AFTERNOON SESSION				
Invited	14.30-15.00	Guillermo Cabrera	Universidad de Conception (Chile)	Unleashing the Power of Transformers for the Analysis of Cosmic Streams	Transformers are deep learning architectures that have shown to reach state-of-the-art performances across various domains. They originally gained attention thanks to their performance in natural language processing tasks. More recently, they have been successfully applied to tasks involving images, tabular data, and time series, among others. In this talk we will review the recent advancements in transformers when applied to the characterization of astronomical light-curves, from task-specific models to foundation models. Transformers have become the new state-of-the-art and will play a key role in analysing cosmic streams in real time from current and next-generation time-domain instruments such as the Vera C. Rubin Observatory and its Legacy Survey of Space and Time (LSST).
Invited	15.00.15.30	Kai Polsterer	Heidelberg Institute for Theoretical Studies - Germany	From Photometric Redshifts to Improved Weather Forecasts: an interdisciplinary view on machine learning in astronomy	The amount, size, and complexity of astronomical data-sets is growing rapidly in the last decades. Now, with new technologies and dedicated survey telescopes, the databases are even growing faster. Besides dealing with poly-structed and complex data, sparse data has become a field of growing scientific interest. By applying technologies from the fields of computer sciences, mathematics, and statistics, astronomical data can be accessed and analyzed more efficiently. A specific field of research in Astroinformatics is the estimation of the redshift of extra-galactic sources, a measure of their distance, by just using sparse photometric observations. Observing the full spectroscopic information that would be necessary to directly measure the redshift, would be too time-consuming. Therefore, building accurate statistical models is a mandatory step, especially when it comes to reflecting the uncertainty of the estimates. Statistics and especially weather forecasting has introduced and utilized proper scoring rules and especially the continuous ranked probability score to characterize the calibration as well as the sharpness of predicted probability density functions. This talk presents what we achieved when using proper scoring rules to train deep neural networks and to evaluate the model estimates. We present how this work led from well calibrated redshift estimates to an improvement in statistical post-processing of weather forecast simulations. The presented work is an example of interdisciplinarity in data-science and how methods can bridge between different fields of application.
	15.30-16.00	*coffee break*			
contributed	16.00-16.20	Joshua Weston	Queen's University Belfast - Ireland	Improvements to the ATLAS Real-Bogus Classifier	We present a Convolutional Neural Network (CNN) for use in the Real-Bogus classification of transient detections made by the Asteroid Terrestrial Impact Last Alert System (ATLAS) and subsequent efforts to improve performance since initial development. As the latest transient surveys increase their scope and depth a greater number of bogus detections will be generated, necessitating improved classification techniques. In such surveys the number of alerts made greatly outstrips the capacity for human scanning, necessitating the use of machine learning aids to reduce the number of false positives presented to annotators. ATLAS has three functioning telescopes each with a different detector and optical performance and here we investigate if training a CNN on data from the individual units improves results compared to a combined training data set. We take a sample of recently annotated data from each of the three operating ATLAS telescopes with ~340,000 real with ~340,000 real (mostly known asteroids) and ~1,030,000 bogus detections per model. Retraining our current CNN architecture with this data we see an improvement to the median False Positive Rate (FPR) from 2.97% to 1.11% with a fixed missed detection rate of 1.00%. Further investigations on the effect of reducing input image size show a negative impact to the false positive rate. Finally architecture adjustments and comparisons to contemporary CNNs indicate our retrained classifier provides an optimal FPR. We conclude that the periodic retraining and readjustment of classification models on survey data can yield significant performance improvements as changes to the data collection process lead to new features in the model input over time.
contributed	16.20-16.40	Sebastian Gomez	HITS - Heidelberg Institute for Theoretical Studies Germany	Unsupervised learning for agnostic knowledge discovery from simulations	Simulations are the best, and often the only approximation to experimental laboratories in astrophysics. However, the complexity and richness of their outputs severely limits the interpretability of their predictions. We describe a new conceptual approach to obtaining useful scientific insights agnostically from a broad range of astrophysical simulations. These methods can be applied to the largest existing simulations and will be essential to solve the extreme data exploration and analysis challenges posed by the next-generation exascale simulations. Our concept is based on applying nonlinear dimensionality reduction methods to efficiently learn compact representations of simulations in a low-dimensional manifold that naturally describes the data's intrinsic variables. The simulation is seamlessly projected onto its intrinsic latent space for interactive inspection, visual interpretation, and quantitative analysis. We present a working prototype of the concept using an Autoencoder trained on real galaxy images from SDSS and simulated galaxies from Illustris to obtain a natural 'Hubble tuning fork' similarity space that can be visualized interactively on the surface of a sphere.

contributed	16.40-17.00	David Sweeney	University of Sydney	Semi-Supervised Learning for Detecting Gravitationally Lensed Quasars	Gravitationally lensed quasars provide us with a fantastic probe for our expanding universe. Independant to our current methods of measuring the Hubble constant, the differing paths of the quasar images through spacetime can allow us to measure the expansion rate. Unfortunately, these objects are exceedingly rare, occurring only for only ~1/10,000 quasars. The challenge is to find these lensed quasars amongst large astronomical data sets. In contrast to previous attempts, which have only made use of numeric data, we perform semi-supervised classification based on images of quasars. These images are low resolution and noisy, but are enough for experienced astronomers to perform classification. Using virtual adversarial training to take advantage of millions of unlabelled images, we develop a classifier which achieves an F1 score of 0.49 - an extremely impressive result in this domain. Predictions made by this classifier are already being used to select candidates for telescopes around the world.Gravitationally lensed quasars are important objects in astronomy for probing the universe. Unfortunately, these objects are exceedingly rare, occurring only for only ~1/10,000 quasars. The challenge is to find these lensed quasars amongst large astronomical data sets. In contrast to previous attempts, which have only made use of numeric data, we perform semi-supervised classification based on images of quasars. These images are low resolution and noisy, but are enough for experienced astronomers to perform classification. Using virtual adversarial training to take advantage of millions of unlabelled images, we develop a classifier which achieves an F1 score of 0.49 - an extremely impressive result in this domain. Predictions made by this classifier are already being used to select candidates for telescopes around the world.
Panel discussion	*17.00-18.00*	*tbd.*			
October 3		**MORNING SESSION**			
invited	9.30-10.00	Michelle Lochner	University of the Western Cape (South africa)	Enabling New Discoveries with Machine Learning	Abstract: The next generation of telescopes such as the SKA and the Vera C. Rubin Observatory will produce enormous data sets, far too large for traditional analysis techniques. Machine learning has proven invaluable in handling massive data volumes and automating many tasks traditionally done by human scientists. In this talk, I will explore the use of machine learning for automating the discovery and follow-up of interesting astronomical phenomena. I will share an exciting recent MeerKAT discovery made with machine learning and discuss how the human-machine interface will play a critical role in maximising scientific discovery with automated tools.
contributed	10.00-10:20	Szymon Nakoneczny	Caltech (USA)	Reliable QSO classification in ZTF and KiDS surveys	I will present how machine learning can be reliably applied to QSO classification in photometric and time domain data, what are the limitations, and how the problem should be approached. In the present era of ever larger photometric sky surveys, designing efficient automated methods for selection of specific categories of objects based on photometric data is becoming an urgent quest. My work focuses on applying ML to KiDS and ZTF surveys. In scope of KiDS, I will present a catalog of 343,000 QSOs. The catalog is based on optical ugri and near-infrared ZYJHKs bands, and provides robust photometric redshifts with uncertainties at high magnitude depth of r < 23.5. We paid special attention to testing the resulting catalog with methods external to training data, such as QSO number counts from eBOSS models, GAIA parallax, the auto correlation function, and cross-correlation with CMB lensing. I will show how different testing methods lead to different decisions of model architecture and probability cuts for the resulting catalog. I will also explain how ML methods can be successfully applied to observations fainter than the training data known from spectroscopy. It includes building inference subsets using high dimensional t-SNE visualisations, and properly adapting bias vs variance tradeoff with tests against the problem of extrapolation. Our success of the extrapolation challenges the way that models are optimised and applied at the faint data end. As part of the ZTF survey, I will present a study of AGN variability, with the main goal of creating a catalog of AGNs to look for gravitational wave candidates from the LIGO O4. ZTF is one of the largest time domain surveys up to date, and allows to study still not well described AGN variability, which is also crucial for AGN classification in photometric data. I will present initial results of classification using transformer based architecture inspired by natural language models. The model was pretrained on ZTF light curves (ASTROMER, Donoso-Oliva, 2023), and we fine tune it in accordance with machine learning good practices. I will present improvement of variability based classification over standard color-color techniques, as well as improvement of deep learning over non deep learning approaches, especially on the faint data end.
contributed	10.20-10.40	Lara Pallas-Quintela	University of A Coruna (UDC)	SOM Maps for outlier classification in Gaia	During 2022 the Gaia space mission published its 3rd data release (GDR3), being the first time that their spectra came to light to the scientific community. As the aim of this mission is to create the most complete 3D map of the Milky Way, some of these spectra may not match the standards defined in the literature, either because they were never observed or due to side effects when gathering data. To tackle this challenge our research group leads a working package called Outlier Analysis (OA), devoted to performing unsupervised classification, using Self-Organizing Maps (SOM), on those BP/RP spectra which have a low probability of membership to standard astronomical classes (classification outliers), allowing further analysis. As we are dealing with a big data problem, one of our objectives is to speed up the algorithm's execution time. Consequently, we developed an improved version of SOM, called Fast-SOM. In the base SOM version, to compute the winner neuron where a source should lay on, every neuron in the map is traversed. With Fast-SOM, from a certain iteration on, we perform such calculus only at the neighbour neurons. Hereby, we can speedup up to 60% the training process with a minimum precision loss (about 5%). Being our algorithm part of the official software of the mission, at GDR3 was the first time OA results were published, processing 56 million sources. With this milestone, we could demonstrate the performance of the algorithm by making a statistical analysis, where 54% of the sources were placed in good quality neurons. This means that we can assign them a spectral type with a high confidence level, thus allowing those objects not to be considered as outliers for any further scientific application. Due to these encouraging results, in the future we want to go one step further by performing unsupervised classification of the whole Gaia survey (200 million sources) just by tuning the algorithm used at OA. Since training a SOM map with the whole survey would be very costly, we want to train several SOM Maps with reference data and, afterwards, try to classify each GDR4 source in one of those maps. On the one hand, we would not lose the SOM morphological classification and, on the other, it would be much faster than creating a SOM Map with the whole survey, which is expected to contain 2 billion objects.
	10.40-11.20	*Coffee break*			

invited	11.20-11.50	Michelle Ntampaka	Space Telescope Science Institute	The Importance of Being Interpretable: ML as a Partner in Cosmological Discovery	ML could play a crucial role in the next decade of cosmology, leading to transformative discoveries from astronomy's rich, upcoming survey data. While ML has historically been touted as a black box that can generate order-of-magnitude improvements at the cost of interpretability, this does not need to be the case – modern techniques are making it possible to develop ML tools that improve results while still being understandable and leading to physical discoveries. In this talk, I will describe understandable models for interpreting cosmological large scale structure. I will show examples of how machine learning can be used, not just as a tool for getting "better" results at the expense of understanding, but as a partner that can point us toward physical discovery.
Invited	11.50-12.20	Emiliano Merlin	INAF- Rome	From jwst to euclid: new algorithms for the study of the first galaxies	Just-launched space missions like JWST and Euclid have just been launched and are already revolutionising our understanding of many physical phenomena in the local and the distant Universe. In this talk I will concentrate on the field where the synergy between these two missions is the strongest: the search and study of the evolution of galaxies, starting from the first that appeared shortly after the Big Bang. After a short introduction of the current status of this research field, I will discuss the challenges in the interpretation and analysis of the data that are obtained with these new facilities. The large amount of data demand automated procedures for their analysis, and the subtle systematics that may easily affect or bias the results require the development of AI or other advanced techniques to be avoided. I will briefly describe a few test cases where the adoption of advanced algorithms may improve the reliability and accuracy of the analysis.
Contributed	12.20-12.40	Zehao Jin	New York University Abu Dhabi	Discovering Black Hole Mass Scaling Relation with Symbolic Regression	Our knowledge of supermassive black holes (SMBHs) and their relation to their host galaxies is still limited, and there are only around 150 SMBHs that have their masses directly measured and confirmed. Better black hole mass scaling relations will help us reveal the physics of black holes, as well as predict black hole masses that are not yet measured. Here, we apply symbolic regression package PySR (Cranmer 2023) to those directly-measured black hole masses and host galaxy properties, and find a collection of higher-dimensional (N-D) black hole mass scaling relations. These N-D black hole mass scaling relations have scatter smaller than any of the existing black hole mass scaling relations. One of the best among them is log(MBH/M⊙)= 2.85 log (σ0/189) + 1.16 log((B/T)/0.437) − 0.33 log (ρsoi/601) + 8.20, with an intrinsic scatter of ε=0.083 dex and RMSE=0.27 dex in the log(MBH/M⊙) direction, significantly lower than ε ~ 0.3 dex and RMSE = 0.46 dex for the M-σ relation. We offer a set of N-D black hole mass scaling relations with such low scatter to inspire black hole physics and test black hole models implemented in simulations. On the other hand, we also provide a set of black hole mass scaling relations that can be easily used to make black hole mass predictions, such that black hole mass can be predicted in its best possible precision when only a limited number of host galaxy properties are known. Acknowledgement: This work is a part of a series of work done in NYUAD by Zehao Jin and Ben Davis (also resigtered for this conference and sumbitted an abstract) on symbolic regression powered black hole mass scaling relations. The organizers can potentially merge the two work into one talk.
contributed	12.40 -13.00	Nikhel Alias Rahul Gupta	CSIRO Space & Astronomy (Australia)	Novel Computer Vision Applications for Radio Astronomy with ASKAP	In this presentation, I will showcase novel applications of computer vision algorithms and data methodologies that we are currently developing. These advancements are aimed at addressing the significant Big Data challenges associated with ongoing surveys like the Evolutionary Map of Universe (EMU) conducted using the Australian Square Kilometer Array Pathfinder (ASKAP) telescope. Specifically, I will talk about the applications of self-supervised algorithms for discovering new radio sources such as Odd Radio Circles. Additionally, I will focus on the weakly-supervised deep learning techniques for detecting extended radio sources, along with the integration of novel semi/fully supervised object detection methods to identify, categorize and group complex radio galaxies with multiple components. Lastly, I will demonstrate the data products generated through these novel computer vision methods, including value-added catalogues of radio galaxies for astronomical and cosmological studies.
	13.00-14.00	*Lunch break*			

AFTERNOON SESSION

invited	14.00-14.30	Nicola Napolitano	Sun Yan Tse University (China) and UNINA	Machine Learning the Universe with upcoming Large Sky Surveys	The emerging "tensions" in the measurement of some critical cosmological parameters and discrepancies between observations in galaxies and cosmological simulations suggest that the "consensus" cosmological model, founded on the existence of two dark components, the Dark Matter (DM) and the Dark Energy (DE), can be either incomplete or incorrect. The solutions to these "cracks" can reside in the understanding of the connection between the dark components, especially DM, and the baryons, i.e. the regular matter (stars and gas) that constitutes the visible part of galaxies. However, a more complex set of cosmological parameters, or some missing physics, including the nature of dark matter, e.g. some form of warm or self-interacting DM, cannot be excluded. I introduce a project aimed at exploiting data from the major next-generation imaging and spectroscopic surveys, using Machine Learning (ML) tools to 1) fast and accurately extract parameters of billion galaxy samples, and 2) use these huge datasets to constrain cosmology and galaxy formation models. For this last objective, I will present a preliminary application to galaxy cluster simulation data.
invited	14.30-15.00	Ashley Villar	Harvard (USA)	Rapid Inference for Extragalactic Transients in the Era of LSST	Soon, the Legacy Survey of Space and Time (LSST) will commence and drive the discovery rate of extragalactic transients to millions per year. Much work has been done to identify these transients, to classify transients in realtime, and to aid researchers in deciding how to optimize observational resources to characterize these events. Here, I will discuss recent developments in inference (i.e., extracting astrophysical information) for such events. I will focus on the use of emulators to approximate complex physical simulations, and simulation-based inference techniques to accelerate Bayesian inference.

contributed	15.00-15-20	Nicolò Parmiggiani	INAF/OAS Bologna	Deep learning techniques to analyse high-energy data of the AGILE space mission.	AGILE is a high-energy astrophysics space mission launched in 2007, and as of the time of this writing, it remains operational. The AGILE payload comprises the Silicon Tracker (ST), the SuperAGILE X-ray detector, the CsI(Tl) Mini-Calorimeter (MCAL), and an AntiCoincidence System (ACS). The ST, MCAL, and ACS combination forms the Gamma-Ray Imaging Detector (GRID). Alongside scientific telemetry, these detectors generate time series data in the form of count rates or ratemeters (RM), with a time resolution on the order of one second. This contribution presents Deep Learning models developed over the past few years to analyze sky maps and time series data acquired by AGILE detectors. These models aim to enhance the scientific output of the AGILE space mission to improve data analysis of real-time analysis to detect gamma-ray transients and explore AGILE's data accumulated over more than 16 years of observations. The first method developed is a 2D Convolutional Neural Network (CNN) designed to detect Gamma-Ray Bursts (GRBs) in AGILE/GRID sky maps. After training the CNN, we conducted a p-value analysis under various background conditions. This analysis yielded thresholds on the CNN output values associated with the statistical significance of GRB detections. We tested the trained DL model using AGILE/GRID sky maps generated from a list of GRBs sourced from Swift/BAT, Fermi/LAT, and Fermi/GBM catalogs. The CNN detected 21 GRBs from this list with a significance of sigma >= 3, significantly improving upon the standard analysis method based on the Li&Ma technique, which identified only two GRBs from the same list. This result underscores the effectiveness of the CNN in detecting GRBs in AGILE/GRID sky maps, and we plan to incorporate this technique into the AGILE real-time analysis system. Upon confirming a GRB detection, an additional CNN model has been developed to localize GRBs in AGILE/GRID sky maps, performing a regression task. The mean localization error obtained is less than 0.8 degrees. We have also devised a method that performs anomaly detection on time series data generated by the AGILE AntiCoincidence System (ACS) to identify GRBs on ratemeters. This technique is based on a 1D CNN autoencoder architecture. We trained the model using a dataset of background-only multivariate time series (MTS) randomly extracted from ACS data. Subsequently, we calculated p-value distributions and evaluated the trained DL model using a list of GRBs obtained from the GRB Web catalog. The DL model detected 72 GRBs with a significance of sigma >= 3, 15 of which were detected by AGILE for the first time. This method is also applied to MCAL's ratemeters, with this model detecting 138 GRBs with sigma >= 3. Furthermore, we applied machine learning to predict the background values of the ACS based on the orbital parameters of the AGILE satellite, including latitude, longitude, and satellite orientation. This method employs a regression model and is currently in development. Encouraged by promising results, we plan to implement an anomaly detection algorithm that leverages the disparities between predicted backgrounds and actual values in the time series to identify GRBs. The DL models presented here outperform classical analysis methods in various contexts, yielding novel scientific insights. These promising suggest further exploring DL techniques applied to astrophysical data in ground- and space-based high-energy observatories. Lastly, we developed a Quantum Deep Learning (QDL) version of some of the aforementioned methods and compared the results with those obtained using classical DL models. The comparison reveals that the QDL model can achieve nearly identical performance with fewer parameters and a smaller training dataset.
contributed	15.20-15.40	Giuseppe Angora	University of Ferrara	Cluster member detection with Faster Region Convolution Neural Networks in HST Galaxy Clusters	In the current era of big data, developing methods capable of autonomously extracting information from vast multi-dimensional datasets plays a pivotal role. This study delves into the potential of region-based architectures for detecting cluster members (CLMs) within galaxy clusters. The primary objective is to disentangle CLMs from foreground and background sources using Hubble Space Telescope (HST) images synergized with extensive spectroscopic surveys (CLASH-VLT and MUSE). Identifying CLMs allows us to study galaxy properties within dense environments and to map the overall cluster mass distributions by also testing cosmological models. I will present the results achieved by an adapted version of the Faster Region Convolutional Neural Network (FRCNN) architecture, which incorporates an Attention Module into its Region of Interest layer. The training procedure of the FRCNN includes direct processing of multi-band HST ACS Field of View, sized at ~2 arcmin. This approach bypasses the conventional practice of extracting source-centered cutouts. The dataset comprises 15 clusters within a redshift range of $0.2 < z < 0.6$, including ~3800 spectroscopic sources. The implemented FRCNN attains a trade-off between purity and completeness of ~85%, coupled with an average Intersection over Union metric of ~0.9. Furthermore, this approach is extended beyond HST to Euclid mock data and VST imaging.
	15.40-16.10	*coffee break*			
Contributed	16.10-16.30	Grant Merz	University of Illinois Urbana Champaign (USA)		The next generation of wide-field deep astronomical surveys will deliver unprecedented amounts of images through the 2020s and beyond. As both the sensitivity and depth of observations increase, larger numbers of overlapping, or blended sources will be detected. Another observational challenge lies in classifying faint point-like sources as stars or galaxies. Both of these realities can lead to measurement biases that contaminate key astronomical inferences. Efficient deblending techniques in particular are a necessity and have been recognized as a high priority. We implement new deep learning models available through Facebook AI Research's Detectron2 repository to perform the simultaneous tasks of object identification, deblending, and classification on large image cutouts of multi-band coadds from the Hyper Suprime-Cam (HSC) Subaru Strategic Program Data Release 3 Deep and UltraDeep fields. We use existing detection/deblending codes and classification methods to train a variety of deep neural network architectures, including state-of-the-art transformer models. We measure the classification and detection/deblending capabilities of different models and attempt to mitigate biases due to label generation. Once trained, we find that transformer-based models outperform traditional convolutional deep neural networks and are more robust to different contrast scalings. Transformer networks are able to detect and deblend objects closely matching the ground truth, achieving a median bounding box Intersection over Union (IOU) of 0.994. Using high quality class labels from the Hubble Space Telescope, we find that the best-performing networks can achieve near 100% completeness and purity of galaxy selection across the whole test sample and above 60% completeness and 80% purity of stars out to HSC i-band magnitudes of 25 mag. Our code, DeepDISC is publicly available at https://github.com/grantmerz/deepdisc. This framework can be extended to other deep surveys such as the Legacy Survey of Space and Time or Roman to enable fast source detection and measurement.

contributed	16.30-16.50	Martin Topinka	INAF/OAS Bologna	CADET	The study of jet-inflated X-ray cavities provides a powerful insight into the energetics of hot galactic atmospheres and radio-mechanical AGN feedback. By estimating volumes of X-ray cavities, their total energy and the corresponding mechanical jet power required for their inflation can be derived. Properly estimating their total extent is, however, non-trivial, prone to biases, and has so far been done manually by scientists. We present a novel and automatized machine-learning pipeline called Cavity Detection Tool (CADET), developed to detect and estimate the sizes of X-ray cavities from raw Chandra images. The pipeline consists of a convolutional neural network trained for producing pixel-wise cavity predictions and a DBSCAN clustering algorithm, which decomposes the predictions into individual cavities. The convolutional network was trained using mock observations of early-type galaxies simulated to resemble real noisy Chandra-like images. The density distribution of simulated galaxies was approximated using a 3D single or double beta-model into which we insert either one or zero pairs of ellipsoidal cavities. Parameters of the simulated galaxy models and the inserted cavities were generated based on an analysis of 70 nearby early-type galaxies, 33 of which contained one or more pairs of X-ray cavities. The network's performance has been tested on simulated data obtaining an average cavity volume error of 20% at a 90% true-positive rate. For mock images without any X-ray cavities inserted, we obtained a 5% false-positive rate. When applied to real Chandra images, the pipeline was able to recover 91 out of 100 previously known X-ray cavities in nearby early-type galaxies and all 14 cavities in more distant galaxy clusters. Besides that, the CADET pipeline discovered 6 new high-significance cavity pairs in atmospheres of early-type galaxies and galaxy clusters and a number of potential cavity candidates.
Panel discussion	*16.50-17.50*	tbd.			
October 4	**MORNING SESSION**				
invited	9.30-10.00	Gennaro Zanfardino	Virtualitics (USA)	High-Dimensional Data Visualization and AI-Enhanced Exploratory Data Analysis in a Commercial Software Solution	The talk will cover the innovative techniques for exploring and analyzing high-dimensional data within the context of a commercial visualization software. I will start with the challenges of rendering network graphs, scatter plots, and a multitude of other data representations on both desktop and VR platforms, while aiming to offer ergonomic user interaction and visualization. I will continue by introducing the AI-driven analytical tools that the proposed solution provides for identifying and analyzing complex network structures and data patterns. The presented methods are versatile and capable of automatically extracting insights from diverse multi-dimensional datasets—ranging from categorical and numerical data to natural language data—across a wide spectrum of domains; the versatility comes with inevitable trade-offs that I hope will ignite discussions on how to effectively mitigate these challenges, especially at scale.
contributed	10.00-10.20	Fernando Caro	INAF - OAR	Hybrid approach to improve deblending performance using machine learning along with traditional tools	Novel technical challenges derived from the processing of astronomical data are constantly arising with the arrival of the images from the latest space missions that have recently been launched, such as JWST or Euclid, where, for instance, the enhanced depths achieved might pose unexpected issues for tasks like deblending. Although, multiple machine learning models have already been developed to improve the performance of deblending, there are still many reasons that prevent them from being the solution ultimately chosen due to the lack of proper training sets, hardware limitations, reproducibility, among many others. Considering this context, we have explored a hybrid approach where deep learning is used to find the best configuration parameters that allow traditional deblending tools, such as SExtractor, to produce improved segmentation maps considering the specific characteristics of every group of blended sources
contributed	10.20-10.40	Mohammad Hashem Faezi	University of Groningen		Multi-band imaging is essential in astronomy for studying a diverse array of celestial objects, including stars, galaxies, planetary nebulae, and supernovae. Astronomers rely on multi-band imaging to investigate the evolution of these objects, monitoring changes in their brightness and color over time. Furthermore, multi-band imaging plays a pivotal role in identifying and characterizing exoplanets, enabling measurements of their size, temperature, and atmospheric composition. Existing tools for astronomical source segmentation are usually limited to single-band image processing, primarily due to the intricate interplay of information across different bands. To surmount this obstacle, our method, referred to as MMTO, leverages the extraction of significant nodes from multiple max-tree data structures, subsequently extending them into a topologically ordered graph arrangement. Within this framework, we investigate correlations between cross-band max-tree nodes and facilitate the merging of semantically similar components. This is accomplished through the utilization of cosine similarity metrics on feature vectors, which encapsulate gray-level features corresponding to sources. In evaluating the effectiveness of our approach, we conducted comprehensive assessments employing complex simulated image datasets with known ground truth. Our findings show the method's capacity to accurately pinpoint related sources across diverse bands, surpassing the segmentation accuracy achieved by a state-of-the-art component-graph-based method. Additionally, a thorough examination of time complexity reveals a considerable enhancement in processing speed, enabling our method to handle larger image datasets efficiently.
contributed	10.40-11.00	Koketso Mohale	University of the Western Cape (South Africa)	Representation Learning for Unsupervised Machine Learning in Astronomy	The next generation of telescopes such as the Square Kilometre Array (SKA) are expected to produce enormous amounts of high-resolution data. Machine learning, being data driven, is increasingly being applied in astronomy. Unsupervised Learning, a sub-field of machine learning, is a more powerful approach for dealing with large and unlabeled datasets. The two most important unsupervised learning tasks are anomaly detection (detecting anomalous objects) and clustering (detecting groups of similar objects). However these algorithms often fail to work with high dimensional datasets, such as image datasets. This leads to a need for algorithms that reduce the dimensionality of resolved datasets, thereby obtaining what are called representations. We explore the use of deep learning techniques such as transfer learning and self-supervised learning to obtain deep representations of optical morphologies. We apply clustering on this representation space and show that meaningful classes can be obtained in a completely automated way.
	11.00-11.30	*Coffee break*			

Invited	11.30-12.00	Johann Knapen	Institute of Astrophysics of the Canary Islands (Spain)	Training on the interface of astronomy and computer science	Modern developments in fields like astronomy imply the need for advanced computational approaches to handle large data sets. However, most recent and current PhDs in astronomy have received little or no formal training in important areas including computer science, computational methods, software and algorithm design, or project management. We will report on a number of scientific advances on the interface of extragalactic astronomy and computer science that have resulted from our EU-funded Initital Training Network SUNDIAL, in which 15 PhD candidates were trained. We will then discuss how these results will be built upon and expanded in a newly approved Marie-Sklodowska Curie Doctoral Network called EDUCADO (Exploring the Deep Universe by Computational Analysis of Data from Observations). We will train 10 Doctoral Candidates in the development of a variety of high-quality methods needed to address the formation of the faintest observed structures in the nearby Universe, including novel object detection algorithms and object recognition and parameter distribution analyses of complex datasets. We aim to detect unprecedented numbers of the faintest observable galaxies from new large-area surveys. We will study the morphology, populations, and distribution of large samples of various classes of dwarf galaxies, compare dwarf galaxy populations and properties across different environments, and confront the results with cosmological models of galaxy formation and evolution. Finally, we will perform detailed simulations and observations of the Milky Way and the Local Group to compare with dwarf galaxies in other environments. We aim for our interdisciplinary training tools, methods and materials to become publicly available.
invited	12.00 - 12.30	Pranav Sharma	United Nations International Computing Centre - UNICC	Navigating Global Policy and Diplomacy through Astroinformatics: Insights from Science20	In the ever-evolving nexus of global policy and technical innovation, the burgeoning field of astroinformatics emerges as a beacon of collaborative potential. As we stand at the intersection of scientific discovery, policy formulation, and international diplomacy, the Science20 (S20) platform serves as a bridge uniting technical expertise from around the world with policy discours. Astroinformatics, at its core, encapsulates the convergence of astronomy, data science, and cutting-edge technologies. Within this context, the recent Science20 engagement titled "Astroinformatics for Sustainable Development" stood as a pivotal milestone. This symposium, held virtually on July 6th and 7th, 2023, delved into the global panorama of astroinformatics and its profound impact on policy and diplomacy. In this talk, I will present the policy discourse that emerged from the Science20 platform highlighting the various views and translatable policy and science diplomacy frameworks.
Panel Discussion	12.30-13.30	*Astroinformatics, education and society*		Chair: tbd	
	13.30-14.30	*Lunch break*			

FREE AFTERNOON

SOCIAL DINNER At Restaurant Zi'Teresa

October 5		**MORNING SESSION**			
invited	9.30-10.00	Fabrizia Guglielmetti	ESO Munich - Germany	A BRAIN study to tackle image analysis with artificial intelligence in the ALMA2030 era	An ESO internal ALMA development study, BRAIN is addressing the ill-posed inverse problem of image analysis employing astrostatistics and astroinformatics. These emerging fields of research offer interdisciplinary approaches at the intersection of observational astronomy, statistics, algorithm development, and data science. In this study, we provide evidence of the benefits in employing these approaches to ALMA image analysis for operational and scientific purposes. We show the potentials of two techniques (RESOLVE and DeepFocus), applied to ALMA calibrated science visibilities. Significant advantages are provided with the potential to improve the quality and completeness of the data products and overall processing time. Both approaches evidence the logical pathway to address the incoming revolution in data analysis dictated by ALMA2030. Moreover, we bring to the community additional products through a new package (ALMASim) to promote advancements in these fields, providing a refined ALMA simulator usable by a large community for training and/or testing new algorithms.
invited	10.00-10.30	Ashish Mahabal	Caltech (USA)	Anomaly detection: overview and application to ZTF	Optical astronomy has become increasingly data rich with billions of sources and hundreds of observations per source. Many methods are being planned to use data-driven ways to classify objects, be they in the Solar System, or variable stars in our Galaxy, or extra-galactic objects. While current surveys and experience over a few decades has prepared us for well understood classes, data covering wider parameter spaces are likely to harbor classes and phenomena not encountered before. There would be in-class extreme outliers, and totally new phenomena. Identifying all kinds of anomalies is crucial to advance our understanding of the cosmos. The most exciting anomalies would be those that reveal our biases and selection effects, and lead us to entire populations unknown so far. We are also prepared for finding a lot of artifacts in the process. A bit like bugs, these could reveal shortcomings in the processing pipelines, and an opportunity for improvement. We will provide a broad summary of anomaly detection and present our work on finding anomalies in ZTF data starting from light curve features and methods like HDBSCAN and isolation forest to look for outliers. We will finish by showing how this connects with the bigger picture, and can be generalized for other surveys.

Contributed	10.30-10.50	Verlon Etsebeth	University of the Western Cape (South Africa)	Astronomaly at Scale: Searching for Anomalies Amongst 4 Million Galaxies	Modern astronomical surveys are already producing vast datasets, with upcoming surveys such as the SKA expected to produce datasets of unprecedented size and richness, increasing the potential for high-impact scientific discovery. However, such volumes of data are impossible to explore without automated means. The development of novel machine-learning-based anomaly detection approaches, such as Astronomaly, has been driven by this possibility and the challenge of exploring such a large number of sources. Astronomaly has the capability to handle various types of data, including images, spectra, or time series data, and can leverage domain knowledge and user preferences to improve detection accuracy and efficiency. For the first time, we test the scalability of Astronomaly by applying it to almost 4 million images of galaxies from the Dark Energy Camera Legacy Survey. We use a trained deep learning algorithm to learn useful representations of the images and pass these to the anomaly detection algorithm Isolation Forest, coupled with Astronomaly's active learning method, to discover interesting sources. We find that data selection criteria have a significant impact on the trade-off between finding rare sources such as strong lenses and introducing artefacts into the dataset. We demonstrate that active learning is required to identify the most interesting sources and reduce artefacts, while anomaly detection methods alone are insufficient. Using Astronomaly, we find 1648 anomalies among the top 2000 sources in the subset after labelling, including 20 strong gravitational lens candidates, 1609 galaxy merger candidates, and 19 previously unidentified sources exhibiting highly unusual morphology. Our results show that by leveraging the human-machine interface, Astronomaly can rapidly identify sources of scientific interest even in large datasets.
	10.50-11.30	*Coffee break*			
Invited	11.30-12.00	Nikos Gianniotis	Heidelberg Institute for Theoretical Studies - Germany	Probabilistic Cross Correlations for Density Estimation	Probabilistic Cross-Correlation for Delay Estimation The Interpolated Cross-Correlation Function (ICCF) has been the workhorse of astronomers when estimating the delay between pairs of lightcurves originating from AGN. We present a probabilistic reformulation of ICCF that enjoys several benefits such as accounting for measurement error, out-of-sample predictions, and most importantly, the capability of delivering a posterior distribution of the delay. Our reformulation views the observed lightcurves in each band as the manifestation of a common latent signal that we model as a sample from a Gaussian process. We demonstrate the advantages of the probabilistic cross-correlation arising from its probabilistic grounding on a number of AGN objects.
invited	12.00-12.30	Michael Wilkinson	RUG - The Netherlands	Connected Morphological Operators in Astronomy: Tools for Object Detection and Pattern Analysis on Vast Data.	Connected morphological filters have seen rapid development theory, algorithms, and applications in many fields of computer vision. More recently, they have found use in astronomical applications, most notably in object detection and pattern analysis. Because these operators work on connected regions in the image in a data-driven way, rather than applying some arbitrary kernel to all pixels equally, parallel algorithms were initially not available. In the past decade, this situation has been amended, and parallel and distributed algorithms capable of handling giga- and tera-scale data sets are available. In this talk I will give an overview of existing tools and algorithms, and new developments in multi-band object detection, dealing with huge data cubes from e.g. LOFAR, exploratory data analysis, and combining these methods with machine-learning tools.
contributed	12.30-12.50	Riccardo Crupi	Università di Udine	Searching for long faint astronomical high energy transients: a data driven approach	HERMES (High Energy Rapid Modular Ensemble of Satellites) pathfinder is an in-orbit demonstration consisting of a constellation of six 3U nano-satellites hosting simple but innovative detectors for the monitoring of cosmic high-energy transients. The main objective of HERMES Pathfinder is to prove that accurate position of high-energy cosmic transients can be obtained using miniaturized hardware. The transient position is obtained by studying the delay time of arrival of the signal to different detectors hosted by nano-satellites on low Earth orbits. To this purpose, particular attention is placed on optimizing the time accuracy, with the goal of reaching an overall accuracy of a fraction of a microsecond. In this context, we need to develop novel tools to fully exploit the future scientific data output of HERMES Pathfinder. In this paper, we introduce a new framework to assess the background count rate of a space-born, high energy detector; a key step towards the identification of faint astrophysical transients. We employ a Neural Network (NN) to estimate the background lightcurves on different timescales. Subsequently, we employ a fast change-point and anomaly detection technique to isolate observation segments where statistically significant excesses in the observed count rate relative to the background estimate exist. We test the new software on archival data from the NASA Fermi Gamma-ray Burst Monitor (GBM), which has a collecting area and background level of the same order of magnitude to those of HERMES Pathfinder. The neural network performances are discussed and analyzed over period of both high and low solar activity. We were able to confirm events in the Fermi/GBM catalog, both Solar Flares and Gamma-Ray Bursts (GRBs), and found events, not present in Fermi/GBM database, that could be attributed to Solar Flares, Terrestrial Gamma-ray Flashes, GRBs, Galactic X-ray flash. Seven of these are selected and analyzed further, providing an estimate of localisation and a tentative classification.
contributed	12.50 -13.10	Ylenia Maruccia	Institute for Space Astrophysics and Planetology (INAF-IAPS)	Characterisation of the Hi-GAL clumps parameter space through experiments of Feature Selection	Star formation studies are benefiting from the huge amount of data made available by recent large-area Galactic plane surveys conducted between 2 µm and 3 mm. The complete characterization of star formation requires combining the information derived from the far-infrared (FIR) to sub-millimetre (sub-mm) range, tracing the very early stages of the process, with the observations in the near-/mid-infrared (NIR/MIR), which allow to trace the later stages of the star formation process, shortly before the appearance of a main sequence star, characterized by the presence of young stellar objects (YSOs). However, the final dataset is often made up of a rich set of heterogeneous and intricate features, and traditional methods of analysis could not be able to uncover hidden patterns and relationships within them. In this perspective, I will present a preliminary study on the characterization of the Hi-GAL (Herschel InfraRed Galactic Plane Survey) clumps parameter space through experiments of Feature Selection. The starting parameter space is our multi-wavelength catalogue of all the star forming clumps present in the Galactic plane region 10° ≤ l ≤ 60°, composed by 6,940 clumps populated by a set of 10,600 YSOs, and containing various types of information: photometric features (i.e. the flux values generated by photometric measurements across different wavelengths and the combination of colours), physical properties (such as clump bolometric luminosities and temperatures, masses, dust temperatures), spatial features (such as sizes and shapes). Moreover, our approach is based on Random Forest, used for data dimensionality reduction by retaining the most important information and for the choice of the most informative features of the star-forming process, with the final aims of (a) obtaining a robust and accurate classification of the Hi-GAL clumps in terms of their evolutionary properties, (b) establishing whether or not a connection exists between the cold material reservoir present in clumps, traced by FIR/sub-mm emission, and the already formed YSOs, precursor of stars. I will discuss results by investigating the outcomes obtained when considering a sub-sample of and the whole parameter space, and making a final comparison between the features identified by our method and those "expected" by theory.
	13.10-14.00	*Lunch break*			
		AFTERNOON SESSION: Quantum Computing			

invited	14.00-14.30	G. Acampora	University of napoli Federico II	Quantum Computational Intelligence	The world of computing is going to shift towards new paradigms able to provide better performance in solving hard problems than classical computation. In this scenario, quantum computing is assuming a key role thanks to the recent technological enhancements achieved by several big companies in developing computational devices based on fundamental principles of quantum mechanics: superposition, entanglement and interference. These computers will be able to yield performance never seen before in several application domains, and the area of artificial intelligence may be the one most affected by this revolution. Indeed, on the one hand, the intrinsic parallelism provided by quantum computers could support the design of efficient algorithms for artificial intelligence such as, for example, the training algorithms of machine learning models, and bio-inspired optimization algorithms; on the other hand, artificial intelligence techniques could be used to reduce the effect of quantum decoherence in quantum computing, and make this type of computation more reliable. This talk aims at introducing the auditors with this new research area and pave the way towards the design of innovative computing infrastructure where both quantum computing and artificial intelligence take a key role in overcoming the performance of conventional approaches.
Invited	14.30.15.00	F. Tacchino	IBM Research- Zurich (Switzerland)	Quantum computing for natural sciences and machine learning applications	Over the last few decades, quantum information processing has emerged as a gateway towards new, powerful approaches to scientific computing. Quantum technologies are nowadays experiencing a rapid development and could lead to effective solutions in different domains including physics, chemistry, and life sciences, as well as optimization, artificial intelligence, and finance. To achieve this goal, noise-resilient quantum algorithms together with error mitigation schemes have been proposed and implemented in hybrid workflows with the aim of improving the synergies between quantum and classical computational platforms. In this talk, I will review the state-of-the-art and recent progress in the field, both in terms of hardware and software, and present a broad spectrum of potential applications, with a focus on natural sciences and machine learning.
Panel Discussion	15.00-15.40	Quantum Computing and astronomy		Chair: Matthew Graham	
	15.40-16.10	*coffee break*			
contributed	16.10-16.30	Benjamin Davis ans Zehao Jin	New York University Abu Dhabi	A Planar Black Hole Mass Relation	Supermassive black holes (SMBHs) are tiny in comparison to the galaxies they inhabit, yet they manage to influence and coevolve along with their hosts. Evidence of this mutual development is observed in the structure and dynamics of galaxies and their correlations with black hole mass (M•). For our study, we focus on relative parameters that are unique to only disk galaxies. As such, we quantify the structure of spiral galaxies via their logarithmic spiral-arm pitch angles (ϕ) and their dynamics through the maximum rotational velocities of their galactic disks (vmax). In the past, we have studied black hole mass scaling relations between M• and ϕ or vmax, separately. Now, we combine the three parameters into a trivariate M•–ϕ–vmax relationship that yields best-in-class accuracy in the prediction of black hole masses in spiral galaxies. We have applied modern machine learning methods to find the best scaling relation, which ideally optimizes both accuracy and simplicity. Symbolic Regression is a sub-field of machine learning that aims to find mathematical expressions that best fit a given set of data. Symbolic Regression searches over equations made of possible selections and combinations of variables, operators, and constants, and judges these equations with a score defined by both accuracy and simplicity. In this work, we adopt the symbolic regression package PySR (Cranmer 2023), which conducts the equation search through a multi-population evolutionary algorithm. The final equation aims to maximize the accuracy and penalize the complexity with a parsimony constant. Our derived fundamental plane represents the best combination of variables and mathematical operations to describe our dataset of directly-measured SMBH masses and their host galaxy parameters. Because most black hole mass scaling relations have been created from samples of the largest SMBHs within the most massive galaxies, they lack certainty when extrapolated to low-mass spiral galaxies. Thus, it is difficult to confidently use existing scaling relations when trying to identify galaxies that might harbor the elusive class of intermediate-mass black holes (IMBHs). Therefore, we offer our novel relationship as an ideal predictor to search for IMBHs and probe the low-mass end of the black hole mass function by utilizing spiral galaxies. Already with rotational velocities widely available for a large population of galaxies and pitch angles readily measurable from uncalibrated images, we expect that the M•–ϕ–vmax fundamental plane will be a useful tool for estimating black hole masses, even at high redshifts.
contributed	16.30-16.50	Giuseppe Dilillo	INAF-IAPS	Detecting astrophysical transients in count time series with Poisson-FOCuS	Gamma-ray burst detection algorithms have remained largely unchanged for over five decades. In this presentation, we explore the potential for making a faster and more sensitive algorithm by leveraging modern anomaly detection techniques. We discuss the conventional approach to gamma-ray burst detection and the application of a novel technique for changepoint detection called Poisson-FOCuS. Poisson-FOCuS enables testing for transients onset over all intervals in a count time series, across all timescales and offsets, efficiently and in real-time. We validate an implementation with automatic background assessment through exponential smoothing, using archival data from Fermi-GBM. Through simulations of lightcurves modeled after real short and long gamma-ray bursts, we demonstrate that the detection power of the same implementation is higher than algorithms designed to emulate the logic of Fermi-GBM and Compton-BATSE, reaching the performances of a brute-force benchmark with oracle information on the true background rate, when not hindered by automatic background assessment. Finally, we show that Poisson-FOCuS can analyze data twice as fast as the best benchmark emulator we could come up with to emulate the logic of Fermi-GBM, using simulated data with different lengths and means.
contributed	16.50-17.10	Demetra de Cicco	University of napoli Federico II	Finding optically variable AGN with machine learning techniques	One of the properties characterizing active galactic nuclei, which allows their identification and selection with respect to other types of astronomical sources, is their variability, which can be observed at all wavelengths. Optical wide-field surveys -such as the Legacy Survey of Space and Time by the Vera C. Rubin Observatory- expected for the very near future will revolutionize astronomy providing us with unprecedented amounts of data, and thus requiring us to gain adequate expertise and methodologies in order to effectively handle them. Optical wide-field surveys -such as the Legacy Survey of Space and Time by the Vera C. Rubin Observatory- expected for the very near future will revolutionize astronomy providing us with unprecedented amounts of data, and thus requiring us to gain adequate expertise and methodologies in order to effectively handle them. The present project is centered on the optimization of AGN selection via their optical variability in wide-field surveys making use of several machine learning techniques. We aimed at testing different models - and hence the corresponding algorithms- to select AGN. The dataset consists of 54 visits of the COSMOS field, obtained with the VLT Survey Telescope (VST) over a 3.3 yr baseline.

contributed	17.10-17.30	Aditya Narendra	Astronomical Observatory of Jagiellonian University (Poland)	Predicting the Redshift of gamma ray loud AGNs using Machine Learning	Active galactic nuclei (AGNs) are very powerful galaxies characterized by extremely bright emissions coming from their central massive black holes. Knowing the redshifts of AGNs provides us with an opportunity to determine their distance to investigate important astrophysical problems, such as the evolution of the early stars and their formation, along with the structure of early galaxies. However, redshift determination is challenging because it requires detailed follow-up of multiwavelength observations, often involving various astronomical facilities. In this presentation I will discuss about the methodology developed by our team, where we apply a powerful machine learning technique called SuperLearner to estimate the redshift of gamma-ray loud AGNs from the Fermi 4th LAT catalog (4LAC). Using the 4LAC's observed properties we train the machine learning model on 1112 AGNs, obtaining a correlation of 75% between the predicted and observed redshift. We also explore the application of an imputation method called Multivariate Imputation by Chained Equations (MICE), using which we impute missing data for 24% of the catalog and proceed to investigate its effects on the redshift estimation. Further, we also explore the application of bias corrections and feature engineering for improving our results. Finally, we provide predicted redshift for 300 BL Lacertae Quasars of the 4LAC using our methodology.
Panel discussion	*17.30-18.30*	*Chair: A. Mahabal*		*What future for astroinformatics*	
October 6		**MORNING SESSION**			
invited	9.30-10.00	Pablo Gomez	European Space Agency	Solving Large-scale Data Challenges with ESA Datalabs	Current and upcoming space science missions will produce petascale data in the coming years. This requires a rethinking of data distribution and processing practices. For example, the Euclid mission will be sending more than 100GB of compressed data to Earth every day. Analysis and processing of data on this scale requires specialized infrastructure and toolchains. Further, providing users with this data locally is not practical due to bandwidth and storage constraints. Thus, a paradigm shift of bringing users' code to the data and providing a computational infrastructure and toolchain around the data is required. The ESA Datalabs platforms is specifically focused on fulfilling this need. It provides a centralized platform with access to data from various missions including the James Webb Space Telescope, Gaia, and others. Pre-setup environments with the necessary toolchains and standard software tools such as JupyterLab are provided and enable data access with minimal overhead. And, with the built-in Science Application Store (SCIAPPS), a streamlined environment is given that allows rapid deployment of desired processing or science exploitation pipelines. In this manner, ESA Datalabs provides an accessible and potent framework for high-performance computing and machine learning applications. While users may upload data, there is no need to for downloading data, thus mitigating the bandwidth burden. As the computational load is handled within the computational infrastructure of ESA Datalabs, high scalability is achieved, and resources can be requisitioned as needed. Finally, the platform-centric approach facilitates direct collaboration on code and data. Currently, the platform is already available to several hundred users, regularly showcased in dedicated workshops and interested users may request access online.
contributed	10.00-10.20	Samira Rezaei	Leiden Observatory - The Netherlands	From Source Detection to Rare Object Detection with AI	With the rise of big data era, the astronomy community have adopted data-driven approaches such as machine learning (ML) algorithms and visualisation techniques into their research. There are several opportunities provided by this mysterious black box set of algorithms that can outperform traditional approaches. The aim of these methodologies is to minimize the need for human interactions, while still providing robust results. This talk would have an interdisciplinary approach and uses knowledge in computer science to advance our understanding of astronomical datasets. The main objectives of this talk can be categorized into followings. First, it provides a pipeline for source detection and characterization to localize the source in observed astronomical images. Beside source detection, the implemented pipeline can remove the observational noise, restore the structure of the celestial sources and predict their properties, such as size and brightness. Then, I will talk about rare galaxy (specifically strong gravitationally lensed systems) classification using ML and the challenges that come with it. Next, I will discuss the challenges when applying such an algorithm to predict the parameters of a strong lens system such as the Einstein radius and source position. Finally, I will talk about model generalization and the possibility to use a developed model for different purposes.
contributed	10.20-10.40	Jingyi Zhang	National Astronomical Observatories, Chinese Academy of Sciences, Beijing, China		Observing Young Stellar objects (YSOs) in stellar clusters and molecular clouds is a common strategy to characterise star forming regions. Their presence attests star formation activity, their spatial distribution within a molecular complex provides clues about its star formation history. The ZTF enables high quality and reliable data products for studying YSOs. Concurrently, the past two decades have seen an explosion of the use of Machine Learning (ML) methods that are also increasingly employed in astronomy. In the present work we first describe our construction of a Self Paced Ensemble (SPE) classifier to improve a widely adopted classification scheme for YSO candidates. The performance of classifier presents that SPE is better than traditional imbalanced learning algorithm for the minority classes. In our work, the SPE classifier of YSOs is rather satisfactory, and the completeness (Recall) of YSOs is enhanced to 91%. As a result, 868,371 ZTF sources are classified into 15 classes by this classifier, which contains 8,210 YSO candidates (YSO prob ≥ 0.70). In order to further identify YSO candidates, these candidates are cross-matched with LAMOST DR9. Finally 833 candidates are observed by LAMOST, among them 379 objects are known YSOs in SIMBAD. For the remaining objects with good-quality LAMOST spectra, we visually check their spectral characteristics, and 238 objects are new confirmed YSOs. These new found YSOs supplement the present YSO sample, and other YSO candidates may be used for follow-up observation, which is useful to characterize YSOs, find more YSOs, and then give a better stellar evolution model in the future. The classified ZTF sources by SPE provide reference to the study of variables and transients. For the new confirmed YSOs, we will utilize the effective temperatures and additional photometric data to build the SEDs of objects in the next work. Then we may determine the masses and ages based on theoretical tracks. In addition, we can calculate the mass accretion rates through the Hα emission lines. Finally we can have a good understanding of the stellar evolution.
contributed	10.40-11.00	Nayyer Raza	McGill University (Canada)	Explaining the predictions of a machine learning classifier for gravitational-wave events	Multi-messenger observations of both the gravitational waves and electromagnetic emission from compact object mergers can give unique insights into the structure of neutron stars, the formation of heavy elements, and the expansion rate of the universe. With the LIGO-Virgo-KAGRA gravitational-wave detectors having recently begun their fourth observing run, it is an exciting time for detecting such events. However, assessing whether to follow up a candidate gravitational-wave event given limited telescope time and resources is challenging; the candidate can be a false alert due to detector glitches, or not have any detectable electromagnetic counterpart even if it is real. To facilitate follow-up decisions, GWSkyNet-Multi was developed as a machine learning model for real-time classification of candidate events, using limited information released in rapid public alerts. While the model is successful, as a deep learning classifier involving convolutional neural networks its inner workings are not transparent and thus its classifications can be hard to interpret. In a recent study, we take a deep dive into explaining GWSkyNet-Multi using novel approaches, including model input tests, to identify the information that is the strongest predictor for GWSkyNet-Multi's output and to understand its misclassifications with an eye towards improving performance. Our perturbation methodology is adaptable to similar traditional models, and can be incorporated into explainability studies for other astronomy machine learning tools to provide deeper insights.
	11.00-11.30	*Coffee break*			

Invited	11.30.12.00	Giuseppe Riccio	INAF- Napoli	An advanced infrastructure for scientific and instrumentation data analysis in Astronomy	in the last decade, Astronomy has entered the big data era, being the scene of the realization of panchromatic surveys, with ground-based and space-borne instruments, characterized by a wide field of view, combined with a very high spatial resolution, capable of acquiring a huge quantity of exceptional quality and deep data. This poses two needs: i) integrating advanced data-driven science methodologies for the automatic exploration of huge data archives; ii) to have efficient short- and long-term monitoring and diagnostics systems, to keep the quality of the observations under control, detecting and limiting anomalies and malfunctions, and facilitating rapid and effective corrections, in order to guarantee the correct maintenance of all components and the good health of scientific data over time. In particular, this is a crucial requirement for space-borne observation systems, both in logistical and economic terms. We present AIDA (Advanced Infrastructure for Data Analysis), a portable and modular web application, designed to provide an efficient and intuitive software framework to support monitoring of data acquisition systems over time, diagnostics and both scientific and engineering data quality analysis, especially suitable for astronomical instruments. Its versatility makes it possible to extend its functionalities, by integrating and customizing monitoring and diagnostics systems, as well as scientific data analysis solutions, including machine/deep learning and data mining techniques and methods. Due to these properties, a specific version of AIDA is already the official monitoring and analysis tool for the ESA Euclid space mission and another one is going to be used for the commissioning of the V. Rubin Telescope.
contributed	12.00-12.20	Cristobal Donoso	Universidad de Concepcion (Chile)	Companion Detection in High-Contrast Images Using Deep Learning	The detection of planets around stars poses significant challenges, especially when faint companions have magnitudes substantially different from the host star. To date, the community has relied on image processing techniques to identify and characterize physical parameters such as position and flux of potential companions. However, while the underlying principles behind the detection of companions are similar, these techniques are often optimized for particular datasets, thereby limiting their applicability to new data. Training algorithms based on such techniques can be resource-intensive in terms of time and space, and they are highly sensitive to initial parameters. Moreover, scalability issues restrict their feasibility to process large sets of images or assess the technique's impact on contrast curves.In this research, we propose a deep learning model using recurrent convolutional neural networks that exhibits remarkable generalization capabilities by training with simulated labels. Our model achieves an average test classification accuracy of approximately 92% when identifying the presence of planets with diverse fluxes and positions around stars. Furthermore, we demonstrate the model's effectiveness in estimating the planetary flux with an R-squared generally smaller than one. We also demonstrate the effectiveness of our model in detecting known companions in real scenarios where planets have been confirmed using classical techniques. These results are based on preliminary experiments, serving as a foundation for future developments. The integration of new datasets and a comprehensive hyperparameter search holds the potential to further enhance our results.
contributed	12.20-12.40	Antonio Ferragamo	Institute of Astrophysics of the Canary Islands & Sapienza Università di Roma	The Three Hundred project: A Machine Lear method to infer clusters of galaxies mass rad from mock Sunyaev-Zel'dovich maps	Our research consists of two closely interlinked projects that highlight the potential of machine learning (ML) in the galaxy clusters study. In the first part of the project, we developed an ML algorithm that utilizes thermal Sunyaev-Zeldovich (SZ) effect maps to infer the radial profiles of total and gas mass in galaxy clusters. Our architecture combines an autoencoder with a random forest to extract information from the maps and perform the final estimation of the mass profiles. To evaluate its effectiveness, we trained and tested the ML algorithm on a comprehensive sample of 73,138 mock SZ maps from THE THREE HUNDRED simulation. Each map in the dataset corresponds to one of the 29 projections encompassing a diverse set of 2,522 galaxy clusters. By analyzing the results, we demonstrate that our model is capable of reconstructing both the gas mass and the total mass profiles, all without relying on any prior assumptions regarding the underlying cluster physics. Our findings reveal that the reconstructed total and gas mass radial profiles exhibit an unbiased nature, with a scatter of approximately 10%. In order to ensure a comprehensive analysis, we deliberately selected galaxy clusters covering a wide mass range, specifically ranging from 10^13.5 to 10^15.5 h^-1 Msun. Additionally, we included clusters in different dynamical states, ranging from highly relaxed to highly disturbed. In the second part of the project, we shifted the focus from reconstructing integrated quantity profiles such as mass to inferring the underlying 2D mass distribution from mock observations of SZ, X-ray and stars using THE THREE HUNDRED simulation. For this aim, we harness the prowess of state-of-the-art deep learning models, lauded for their capacity in image generation. These models are tailored to observed data, such as SZ maps or stellar distribution, in order to generate the most probable 2D matter distribution within our dataset, which encompasses approximately 70 thousand mock maps. Our study underscores the models' adeptness in successfully deducing the 2D matter distribution, reflecting a modest 10% scatter within their mass radial profiles. Additionally, the fidelity of the matter power spectrum is retained as evidenced by the congruence between the generated maps and the ground-truth data extracted from simulations. The logical progression of this project involves the preparation of mock images essential for the training of deep learning models that must comprehensively account for the telescopes' observational effects to accurately emulate authentic real-world observations
Contributed	12.40 -13.00	Nicolò Cibrario	INFN, Università di Torino	Joint machine learning and analytic track reconstruction for X-ray polarimetry with Gas Pixel Detectors	Our study concerns the reconstruction of photoelectron tracks in Gas Pixel Detectors (GPDs) used for astrophysical X-ray polarimetry. The GPD is a detector which exploits the photoelectric effect to measure the polarization of incident photons, by reconstructing the track properties of the emitted photoelectrons. We developed an algorithm which maximizes the performance of Convolutional Neural Networks (CNNs) to predict the impact point of incoming X-rays from the image of the photoelectron track. A very high precision in the reconstruction of the impact point position is achieved thanks to the introduction of an artificial sharpening process of the images. Our CNN is able to correctly interpret the hexagonal structure of the pixels of the track images, thanks to the introduction of hexagonal convolutional blocks. The CNN achieved significant improved performance in the reconstruction of the photon's impact point compared to the state-of-the-art analytic analysis. We find that providing the CNN-predicted impact point as input to the same state-of-the-art analytic analysis improves the modulation factor and naturally mitigates an effect appearing in polarization measurements of bright extended sources known as "polarization leakage".
contributed	13.00-13.20	Jorge Sarrato-Alós	Institute of Astrophysics of the Canary Islands		Accurately determining the mass distribution within galaxies is crucial for understanding their formation and evolution. Previous research has traditionally relied on analytical equations based on the Jeans equation to estimate the enclosed mass with minimum projection effect. In this study, we present a novel approach to predict the enclosed mass within a given radius using a machine learning model trained on line of sight data of high-resolution cosmological hydrodynamical simulations. Our dataset comprises a diverse sample of galaxies spanning a wide range of masses. To train the model, we utilize projected positions and velocities of stars within the galaxies. Multiple training iterations are performed, each with the mass enclosed within a different radius as the target variable. By systematically varying the radius, we identify the optimal value at which the neural network exhibits the highest precision in predicting the enclosed mass. Our results demonstrate the effectiveness of the machine learning-based approach in predicting galaxy mass within a specific radius. The trained model offers a valuable tool for studying galaxy properties, such as mass distribution and gravitational potential, providing insights into the formation and dynamics of galaxies. This work also highlights the utility of machine learning techniques for studying galaxies through line of sight data.
	13.20-13.30	Closing remarks			