

PAPER • OPEN ACCESS

The Spoke 2 of the ICSC National Centre, with a focus on deep learning applications in astroparticle physics and satellite imagery

To cite this article: G. A. Anastasi *et al* 2025 *J. Phys.: Conf. Ser.* **3053** 012031

View the [article online](#) for updates and enhancements.

You may also like

- [Recent outreach activities of the EEE Project](#)
C. Ripoli, M. Abbrescia, C. Avanzini *et al.*
- [GINGER Status Report](#)
Carlo Altucci, Francesco Bajardi, Andrea Basti *et al.*
- [Exploring cosmic rays with educational outreach activities: a long-term study of the correlation between atmospheric conditions and muon counts](#)
D. Miceli, C. Arcaro, S. Canella *et al.*



The Electrochemical Society
Advancing solid state & electrochemical science & technology



249th
ECS Meeting
May 24-28, 2026
Seattle, WA, US
Washington State
Convention Center

Spotlight Your Science

**Submission deadline:
December 5, 2025**

SUBMIT YOUR ABSTRACT

The Spoke 2 of the ICSC National Centre, with a focus on deep learning applications in astroparticle physics and satellite imagery

G. A. Anastasi^{1,2,*}, S. Albergo^{1,2}, N. Pino^{1,2}, G. Piparo²,
S. M. R. Puglia^{1,2}, A. Tricomi^{1,2}

¹Dipartimento di Fisica e Astronomia “E. Majorana”, Università di Catania, Catania, Italy

²Sezione INFN di Catania, Istituto Nazionale di Fisica Nucleare, Catania, Italy

* corresponding author

E-mail: galex.anastasi@dfa.unict.it

Abstract. The National Research Centre (CN) for High Performance Computing, Big Data and Quantum Computing, managed by the ICSC Foundation, has been established under the National Recovery and Resilience Plan as part of the Education and Research Mission. The CN includes an Infrastructure Spoke plus 10 thematic Spokes and, besides building a world-class supercomputing cloud infrastructure, its purpose is to carry out research in computing and high-performance data analysis, identified as strategic areas for any future scientific and technological development. In particular the Spoke 2, dedicated to “Fundamental Research & Space Economy”, addresses the challenge of increasing computational needs in the field of theoretical and experimental collider physics, astroparticle physics and gravitational waves investigation. In this contribution, after a short overview of the CN and of the planned infrastructure, the status and perspectives of the Spoke 2 are presented. Furthermore, two use-cases are illustrated: i) data-driven identification of signals in different experimental apparatuses (in particular a Liquid Argon TPC and a ground array of water-Cherenkov detectors) using self-supervised neural networks; ii) analysis of satellite imagery for the segmentation of wildfire-affected areas, employing supervised deep learning techniques on the data from the Copernicus Sentinel-2 mission and the Copernicus Emergency Management Service.

1 Introduction

The scientific research, particularly at the forefront of knowledge, is increasingly becoming a computing-intensive field. Both current and upcoming experiments exhibit requirements on data processing and storage on par with the leading global entities, and thus necessitate solutions that extend beyond the traditional curriculum of scientists. This trend has characterized the history of computing technologies in collider physics research, but similar needs have now emerged in other scientific domains, like in astro(particle)-physics and gravitational waves, where comparable, if not greater, resource requirements are expected in the near future.

At the same time, major investments have been planned by the US, China, and European Union to support the realization of pre-exascale and exascale digital infrastructures in the coming years. For instance, billions of Euros have been allocated for EuroHPC, an initiative for the development of HPC technologies in Europe, closely aligned with other efforts aimed at enhancing competitiveness in various



domains, such as GAIA-X, a federated and secure sovereign data infrastructure, the European Open Science Cloud (EOSC) and the European Processor Initiative (EPI). Italy aims to make a significant contribution to this effort through its National Recovery and Resilience Plan (PNRR) in the context of the NextGenerationEU program, building an innovative National Center (Centro Nazionale, CN) based on a distributed supercomputing cloud infrastructure.

In this paper, the CN is presented, with a particular emphasis over the initiatives in the “Fundamental Research and Space Economy” task, dedicated to enhancing the scientific capabilities of both current and future experiments. Two use-cases proposed in this context, and currently under study, are then discussed. In both, the aim is to develop machine learning techniques to be employed in the analysis of measurements, either from astroparticle physics experiments or from satellites. In the conclusions, the next steps and future perspectives are addressed.

2 The CN and the Spoke 2

The National Research Centre for High Performance Computing, Big Data and Quantum Computing, managed by the ICSC Foundation, is one of the five National Centres established under the Italian PNRR, as part of the Education and Research Mission coordinated by the Ministry of University and Research. The main goal of this CN is to establish a national infrastructure for cutting-edge research and innovation in the fields of simulations, computing and data analysis. To accomplish such task, it is designed to integrate existing high-performance computing (HPC), high-throughput computing (HTC), big data and network infrastructures with new targeted resources funded within the project.

The CN follows a hub-and-spoke model: the hub is responsible for the validation and management of the research programs, while the spokes and their affiliated institutions develop and implement the chosen activities and use-cases. Both the hub and spokes consist of universities, research institutions, private and public operators, to foster synergy between the scientific community and the industrial sector to the benefit of the national research and production systems. The CN includes one cross spoke, Spoke 0 (“Supercomputing Cloud Infrastructure”), and 10 thematic spokes, as illustrated in Fig. 1.

The infrastructural Spoke 0 aims to consolidate Italy’s network of supercomputing centers and data-intensive centers, currently offering approximately 200 PB of storage capacity and 100,000 CPU cores of computing power. Additionally, the GARR national research network¹ will be appropriately upgraded by enhancing the connectivity of the national backbone and the last mile of major centers to terabit-per-second scales. A crucial concern is to address and overcome regional disparities, particularly in southern Italy; indeed, Spoke 0 plans to upgrade and deploy new additional centres especially in the Southern regions. Additionally, this integration plan comprises the network of INFN data centers, which provides HTC and storage resources to numerous scientific collaborations in particle, nuclear, and astroparticle physics. The INFN data centers have been integrated into the worldwide computing grid since 2003, and therefore brought over twenty years of experience in distributed computing in the CN.

¹<https://www.garr.it/>.

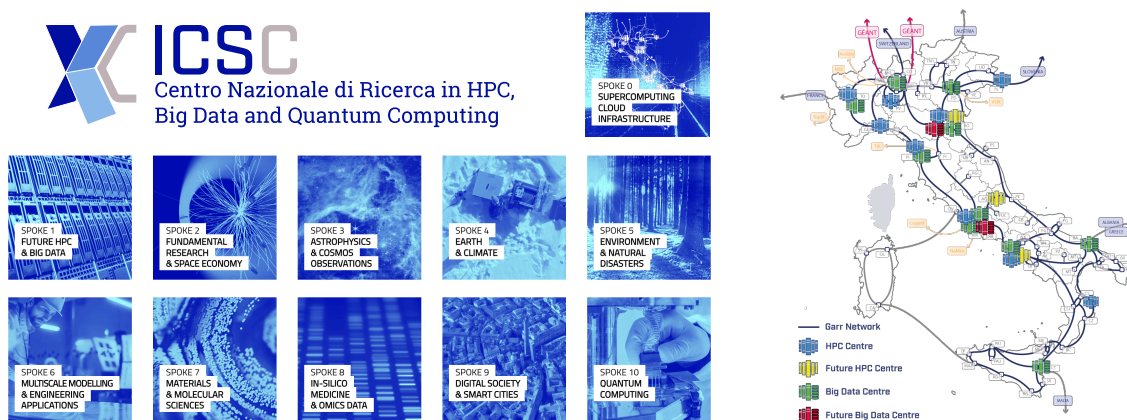


Figure 1: The 11 Spokes forming the CN. The Spoke 0 in particular is dedicated to the implementation, optimization and enhancement of the computing infrastructure, starting from the existing resources connected by the GARR network (on the left). Images from <https://www.supercomputing-icsc.it/>.

2.1 The Spoke 2

The activities within the Spoke 2 (“Fundamental Research & Space Economy”) are framed as research in basic sciences, particularly within the domains of theoretical and experimental physics involving accelerators, space- and ground-based detectors for astroparticle physics and gravitational wave investigations. Over varying time scales, these fields will inevitably encounter challenges related to the scaling and efficiency of computing infrastructures. To address these challenges, the Spoke 2 aims to design, develop, and test solutions suited to both current and next-generation experiments, leveraging the CN infrastructure. This is to be achieved also implementing more efficient strategies to reduce computational costs and minimize the power consumption footprint.

The Spoke 2 activities are organized around six distinct Work Packages:

- WP1: algorithms and computational strategies for theoretical physics;
- WP2: applications for experimental high energy physics (colliders, neutrino experiments, etc.);
- WP3: applications for experimental astroparticle physics and gravitational waves experiments;
- WP4: boosting the computational performance of theoretical and experimental physics algorithms, porting applications to GPUs, FPGAs and heterogeneous architectures;
- WP5: architectural support for theoretical and experimental physics data management on the distributed data-lake CN infrastructure;
- WP6: cross-domain initiatives and Space Economy (more on sect. 4).

Even if the Spoke 2 focuses on advancements in these fundamental research areas, the expectation is to achieve solutions with applications well beyond science, as in other historical examples especially from high-energy physics, including the creation of the World Wide Web in the 1980s and the development of the Grid in the early 2000s. In particular, the tools developed by the more “technical” working packages (WP4-5-6) are conceived to be extendable to other scientific domains within the CN and to the industrial partners in the Spoke. Additionally, the trained personnel will help proliferate HPC methodologies across Italian academic and industrial sectors, fostering comprehensive advancement.

As for the three-year research plan, a first phase (encompassing the first year of the project) has been dedicated to the landscape recognition and the identification of state-of-the-art use-cases. In the current phase, the project staff and newly hired personnel is carrying out the actual development of algorithms and services, which is being documented (at alpha/beta level) and prepared for large-scale testing. In the final phase(s), the outcomes will be validated through test-beds benchmarked to ensure they meet the specified requirements, before being consolidated into reports and white papers.

3 A WP3 use-case : DAIDREAM

In the current era of multi-messenger astronomy, which combines data from various observatories and missions (e.g. gamma- and cosmic-ray observatories, dark matter and energy probes, neutrino telescopes, and gravitational interferometers), new approaches and tools are required to assure the advancement of the research in the field. From simulation to event reduction and reconstruction, the large, interconnected datasets require advanced data sharing and distribution solutions, leveraging fast computing and machine learning techniques. The WP3 is intended to address such tasks, that however encompass several different scientific contexts. For this reason, beyond 7 flagship use-cases considered of particular interest for the goals of the WP3, many ordinary (non-flagship) use-cases have been identified, confirming the diversity and importance of the technological challenges faced by the collaborators.

The project described in this section is included within the area of expertise “Analysis techniques for astroparticles and high energy photons”, and it has been named DAIDREAM (*DA*tA-driven *ID*entification of *R*are *E*vents in *A*stroparticle physics through *M*achine learning techniques). The objective is to employ deep-learning to fully exploit the experimental data, with the implementation and adaptation to our tasks of unsupervised or weakly supervised neural networks. Such techniques have been already employed successfully in fundamental science, for instance in the search for anomalous signatures in high-energy particle physics [1] and for the rejection of noise transients in gravitational waves analysis [2]. We plan the development of this use-case in two distinct (yet contiguous) experimental settings.

A first application was identified in the context of direct searches for Weakly Interacting Massive Particles (WIMPs), among the most promising candidates for dark matter, which may be detected by looking for scattering processes with ordinary matter at the weak scale. The DarkSide program, with the future DarkSide-20k experiment [3], aims at WIMPs detection using a dual-phase (liquid and vapor) Liquid Argon Time Projection Chamber (LAr TPC), where both the prompt argon deexcitation

scintillation light (S1) and the delayed ionization signal (S2) are recorded using a matrix of silicon photomultipliers (SiPM). Within the DarkSide program, an R&D project called *Recoil Directionality* (ReD) experiment [4] has collected a large amount of data to perform a systematic study on nuclear recoils down to very low energies (2-5 keV) in a miniaturized version ($5 \times 5 \times 6 \text{ cm}^3$) of the Darkside-20k TPC. Such reactions are produced by the elastic scattering in argon of neutrons from a ^{252}Cf source, with BaF_2 detectors to tag fission events and plastic scintillators to measure the scattering angle. However, at low recoil energies the scintillation signal (S1) is no longer observable and a large fraction of measurements shows only ionization signals (S2). The study of S2-only events requires a more effective reduction of the background, dominated by the ^{39}Ar and ^{85}Kr decays in the liquid argon bulk, by gamma and X-rays from radioactive isotopes in the detector components, and by electronic recoil processes which cannot be rejected via pulse shape discrimination techniques [5]. For these reasons, the ReD TPC acquisition is performed in slave mode, and the S2-only signals are selected offline by searching for events with a time-of-flight (measured between the BaF_2 detectors and the plastic scintillators) compatible with the propagation of a neutron. State-of-the-art, self-supervised artificial intelligence methods can allow to improve the background rejection in absence of the S1 signal, in order to identify recoil events at low energies, as described in the next paragraph (sect. 3.1).

Another feasible application concerns the field of ultra-high energy cosmic rays (UHECR), i.e. protons and ionized nuclei reaching the Earth with energies above $\sim 10^{17}$ eV. Such particles are studied through the observation of extensive air showers (EAS), cascades of secondary particles created in the interaction of a primary cosmic ray with the atmosphere. In this context, the detection of rare or anomalous events that could be induced by non-hadronic primaries, such as UHE photons and neutrinos, would open a new window on physics processes beyond the Standard Model, or related to dark matter. The largest experiment built for the measurement of UHECRs is the Pierre Auger Observatory [6], that was designed as a hybrid detector and comprises: an array of 1600 water-Cherenkov detectors (WCDs) arranged in a 1.5 km triangular grid over an area of $\sim 3000 \text{ km}^2$ (Surface Detector, SD); 24 fluorescence telescopes overlooking the SD from four buildings located at the edges of the array (Fluorescence Detector, FD). Over the last 15 years, the Observatory has accumulated the world's largest exposure to UHECR, but no UHE γ or ν candidates have been unambiguously identified so far [7]. Such studies, given the reduced duty-cycle of the FD ($\sim 15\%$), are performed through multivariate analyses of the observables acquired by the SD, that only provides a sampling measurement of the shower particles at ground level by recording the spatial and temporal distribution of the signals produced in the WCDs. The interpretation inevitably requires the usage of simulations, where the models for hadronic interactions are extrapolated at energies and physics conditions largely inaccessible at particle accelerators. Consequently, systematic differences with experimental data are observed, as in the case of the muon content (see Ref. [8] and references therein). For these reasons, data-driven applications based on unsupervised or weakly supervised deep learning hold considerable promise to advance these searches. Indeed, one of the use-cases' goal is the search for unusual events (possibly induced by non-hadronic primary particles) in the data of the Pierre Auger Observatory, treating the measurement collected by the SD as a multi-layered map.

The project is under development employing isolated Virtual Machines instantiated in the INFN-Cloud infrastructure² and provided with CPU resources and standard services, in particular Jupyter Notebooks with persistence for the collaborative work among the group members.

3.1 Application on synthetic data resembling the ReD measurements

The method developed to identify S2-only signals produced by low energy nuclear recoils in the ReD TPC is based on an unsupervised deep learning application implemented with an autoencoder architecture. An autoencoder is a neural network formed by an encoder part, that compresses the input data into a space with less dimensions called *latent space* (also known as bottleneck or coding layer), and by a decoder part, that reconstructs the input from the latent space representation. The specific implementation of a (convolutional) autoencoder for this use-case has been made in Keras with Tensorflow backend, and comprises 3 convolutional & average pooling layers with a *ReLU* activation function. A schematic view of the network architecture is depicted in Fig. 2. To train an autoencoder, the network's weights are updated (with a back-propagation algorithm) in the direction that minimizes the differences between the decoder output and the input, which are directly compared; for this reason, such a procedure is also referred to as *self-supervised* learning. In this case, the input are time-series composed by around 10,000 samples per event, and the loss function (which represents the encoding-decoding error) is calculated as the sum of the square differences between the input and output waveforms.

²<https://www.cloud.infn.it/>.

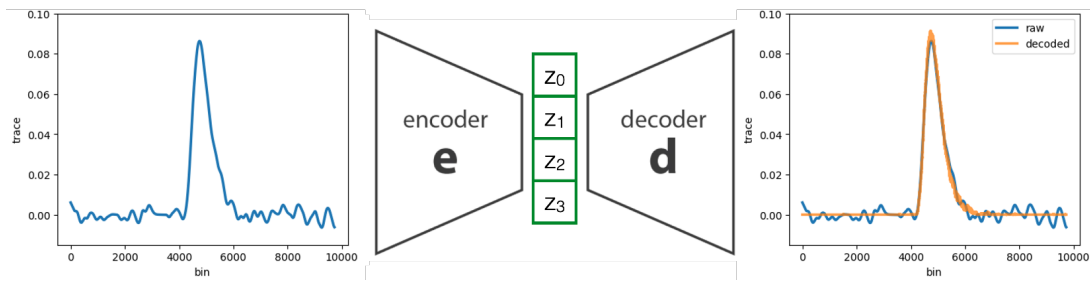


Figure 2: Schematic representation of an autoencoder architecture with a 4 dimensional latent space. On the left, the input waveform in cyan; on the right, the output (i.e. encoded-decoded) trace in orange.

The main advantage of autoencoders is to implicitly highlight the fundamental characteristics of a dataset, as the network is forced to learn a compressed representation while disregarding the background noise and redundancies in the dataset, making this architecture useful in physics tasks like feature extraction, denoising and anomaly detection. Given the specific task of identifying nuclear recoil events over the background but without the requirement of characterizing the corresponding signals, a 4-dimensional latent space has been selected, allowing a direct study of the compressed representation in each of the encoded values, that are indicated as z_i with $i = 0, 1, 2, 3$ in the following.

In a first step of the use-case development, the proposed model has been applied onto synthetic waveforms, which resemble the S2-only events collected by the ReD experiment, in particular the average over the measurements of each single SiPM. Such synthetic traces are composed of a non-gaussian background plus a log-normal shaped signal, extracted uniformly in the decimal logarithm of the integral and re-scaled to have a maximum amplitude between 0 and 1. We generated 10,000 synthetic waveforms and used 6000 for training, 1500 for validation after each epoch, and the remaining 2500 for testing. The network is trained for 100 epochs, using the ADAM optimization algorithm and the *ReduceLROnPlateau* Keras callback to improve the learning process especially during the latest epochs.

The latent space representation is then studied, looking at the distributions of the 4 values in which each synthetic trace is compressed by the network, as in Fig. 3, where the distribution for the latent space dimension z_0 is shown. It can be noticed that the waveforms containing a very small signal (namely with a peak amplitude lower than 0.01, which is comparable with the background noise) are encoded into a specific, small range of values, where an accumulation in the overall distribution is visible. Instead, values of the z_0 parameter outside this accumulation generally correspond to large signals. Such a characterizing behaviour is repeated in the other z_i distributions and can be used to identify events with negligible signals, i.e. not clearly distinguishable from the background noise. In fact, one can define an interval of values for each of the z_i dimensions, and thus delimit a region in the latent space (nicknamed as “garage”) corresponding to the encoding representation of traces with a negligible signal. Then each waveform can be labelled with a simple condition: (a) if the encoded representation falls in the “garage” (meaning that the 4 z_i fall simultaneously in the specified intervals), the event is tagged as noise-only; (b) if not, the event is tagged as signal candidate.

Such a method, when applied on the synthetic dataset, correctly labels 100% of events with large signals, and the fraction of true positives only decreases for peak amplitudes comparable with the accidental pulses in the pedestal. Furthermore, it was verified that the fraction of false positives (intended as events with only background noise but tagged as signals) is $\sim 1\%$. Such results confirms the feasibility of the method, with the promising perspective of a very high efficiency in recognising signal pulses down to the limit represented by the intrinsic noise. The developed application is currently being applied to the ReD data, after the approval of the ReD and Darkside collaborations.

4 A WP6 use-case: AI algorithms for (satellite) imaging reconstruction

The aim of the WP6 is (at least) twofold: on the one hand, to proceed in the adaptation and optimization of widely used software packages (Geant4, FLUKA, etc.) leveraging the CN infrastructure, together with statistical and AI-based data interpretation tools; on the other hand, in the context of the Italian Space Economy Strategy, to develop and deploy applications for accessing and processing satellite data, enabling the creation innovative services such as emergency, risk, and resource planning.

The flagship use-case “*AI algorithms for (satellite) imaging reconstruction*” focuses on deep learning for automatic object detection and classification in aerial and satellite imagery, exploiting in particular

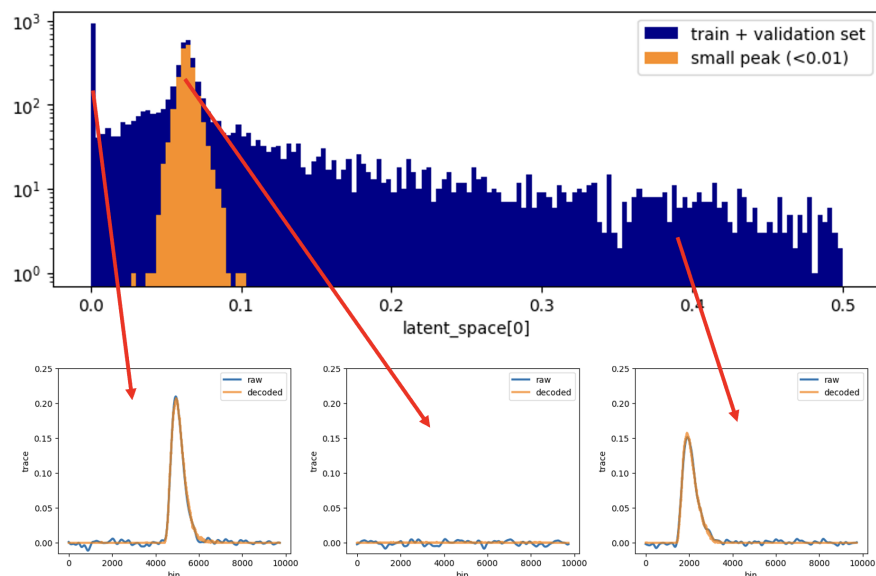


Figure 3: In dark blue, the distribution for z_0 (first dimension of the latent space). An accumulation of values in a range between 0.05 and 0.1 is evident and corresponds to the encoding of waveforms with very small signals. This is emphasised by plotting the distribution (in orange) of the z_0 values obtained for traces with signal amplitude lower than 0.01 (that is comparable with the background noise). Values outside the described range (and in particular $z_0 = 0$) are produced instead from the encoding of traces with large pulses.

the data from the Mirror Copernicus program. Indeed, in the context of Remote Sensing, artificial intelligence algorithms can be effectively used to extract information about land (for instance mode of use or cover evolution), vegetation status, but also extreme events at a global level. Despite the recent advancements, the effectiveness of such methodologies remains yet limited due to the scarcity of available datasets for supervised model training, the need for extensive parameters tuning and the challenges related to inevitable adversarial conditions such as cloud coverage and solar radiance.

The CN offers a unique and promising opportunity to test and improve state-of-the-art computer vision techniques (such as vision transformers), with the goal of demonstrating the capability of deep learning methods to fully grasp the complex and heterogeneous features in the high resolution remote sensing data. After a first phase of study and tracking of the best technologies to be employed, 3 technical methodologies and computational frameworks have emerged:

- **Deterministic Learning algorithm for object identification.** This approach mainly involves the analysis of RGB color values to distinguish a specific class of objects from the surroundings, without using deep learning tools and consequently without requiring large labelled datasets for training. As described in Ref. [9], this method has been already employed successfully in the recognition of photovoltaic panels in aerial images, allowing to estimate the production of green energy expected in a community/region;
- **High-resolution image processing based on spectral indexes for disease detection in vineyards,** in particular to identify the presence of *Flavescence dorée* in imagery collected by Unmanned Aerial Vehicles. The project employs the analysis of RGB indexes to identify spectral signatures unique to the diseased plants, followed by a spatial clustering of the identified disease spots accomplished through the DBSCAN algorithm. This method provides a precise monitoring of the vineyard health and allows targeted interventions on the disease [10].
- **Deep Learning for satellite imagery analysis,** focusing on the detection of diseases in vineyards and on the segmentation of wildfire-affected areas employing the multispectral images acquired by the Sentinel-2 satellites³, which are part of the Copernicus Constellation, coupled with the mapping of burnt sites provided by the Copernicus Emergency Management Service⁴ (CEMS). The

³<https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2>.

⁴<https://emergency.copernicus.eu/>.

applications developed within these projects are convolutional neural networks (CNNs) with Long Short-Term Memory (LSTM) layers to track changes over time (like the progression of diseases or the differences before and after a wildfire) as described in the next section.

4.1 Applications on the Sentinel-2 imagery for the segmentation of wildfire-affected areas

Given the high risk and enormous damage caused periodically by wildland fires, a multitude of systems have been developed over the years for detection, mapping of the affected areas and of the damage severity, and possibly prediction of the fire outbreak and spread. In more recent times, machine learning and in particular deep learning techniques have been employed for these tasks with very promising results (see Ref. [11] for a review) also thanks to the technological advancements in (ground) sensors, cameras, drones, and satellites. Indeed, the goal of this use-case is the automatic detection and delimitation of burnt areas in the images acquired by the Sentinel-2 satellites, using deep neural networks trained with supervised learning procedures on the historical wildfire activation maps from the CEMS. The chosen architecture is the convolutional U-Net, a popular encoder–decoder network used for image segmentation, but with the addition of LSTM layers: the convolutional part extracts the features essential for the identification of the burnt surfaces, while the LSTM part emphasizes the temporal evolution.

A preliminary dataset containing 23 areas of interest at 3 different times (between 30 and 10 days before the wildfire, during the event, and between 10 and 30 days after) has been built, downloading each image with a granularity of (512×512) pixels for each of the 12 spectral bands available. The features selected as input are 10 spectral bands and 9 vegetation indices calculated combining different spectral bands to enhance the presence of fire-affected surfaces⁵. The model is implemented in Keras (with Tensorflow backend) and trained for 50 epochs, using the ADAM optimizer and a custom loss composed of a Dice term, a Jaccard term and a standard cross-entropy loss; 7 images are used for validation. The results are generally very promising, as can be seen in Fig. 4 for one of the areas.

In order to streamline the download and analysis of the Sentinel-2 imagery, a custom python library, based on the Sentinel-Hub API⁶, is being developed. At the moment, it includes 4 modules: one dedicated to the download of images, currently implemented for Sentinel2-L2A products only; a module for the data manipulation, which produces the maps for the single spectral bands and vegetation indexes in TIFF and numpy-native formats, also combining the wildfire information from the CEMS data; a module to visualize the downloaded and processed maps in standard formats (PDF, PNG, etc.); a module to handle the (pre-)processing of the images for the usage in deep learning applications, to facilitate experimentation, optimization and exportation of the implemented tools. Although the library is still under development (in particular the last module), it is already in active use and will form the basis of the open-source repository to be made available by the end of the project.

The described activities are planned to scale-up when the (CPU and GPU) computational resources provided by the ICSC National Center will become fully available, a crucial step for handling large volumes of information with an high level of complexity, and anyway necessary to develop solution apt to meet real-world demands.

⁵<https://www.indexdatabase.de/>.

⁶<https://docs.sentinel-hub.com/api/latest/>.

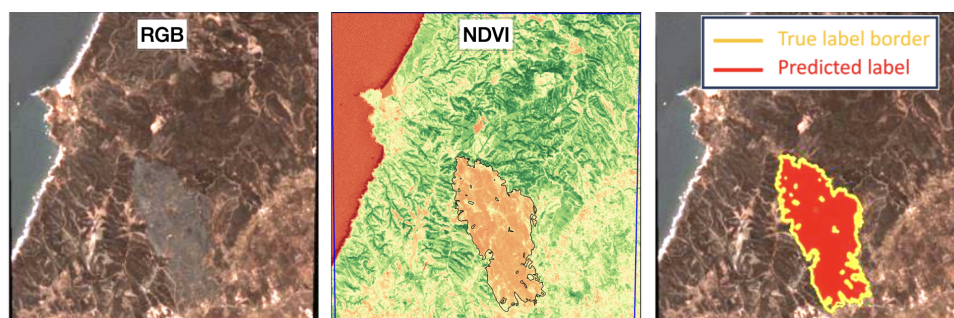


Figure 4: Example forest fire event, started on 2020-06-19 in Algarve, Portugal (entry EMSR443 in the List of CEMS Rapid Mapping Activations). On the left, the post-fire RGB image; at the center, the corresponding visualization for the *Normalized Difference Vegetation Index* (NDVI), together with the border (in black) of the burnt surfaces as supplied by the CEMS service; on the right, the output of the trained network.

5 Conclusions and outlook

In this paper, the purposes and the general structure of the National Centre for HPC, Big Data and Quantum Computing have been discussed, with a focus on specific deep learning applications developed within the Working Packages 3 and 6 of the Spoke 2.

In the context of the use-case DAIDREAM, the next steps involve an extension of the developed tools to exploit the complete information contained in the ReD data, i.e. the waveforms collected by each single SiPM instead of the global average trace. Moreover, the same techniques are going to be applied to the measurements of extensive air-showers at ground level, starting from a preliminary analysis of the Auger Open Data [12], a public dataset comprising more than 25,000 events with primary energy above $\sim 2.5 \times 10^{18}$ eV released by the Pierre Auger Collaboration. In a future perspective, the described methodologies for the detection of rare events in astroparticle physics are planned to be applied in the analysis of data from the ReD extension (ReD plus program, already funded) and from AugerPrime, the Pierre Auger Observatory upgrade [13]. In addition, such tools may also be adapted to other experiments based on noble liquid scintillators like Xenon detectors.

Regarding the segmentation of wildfire-affected areas, noticeable improvements are expected by enlarging the training dataset: the download and pre-processing of the available areas of interest (up to around 120) with a higher granularity (up to 2048×2048 pixels) is currently underway. The training phase can also be improved by introducing data augmentation tools and a better management of the cloud coverage. Following the successful delimitation of burnt surfaces, interesting objectives are the severity estimation and the prediction of wildfires, the latter being a much more complex task that will most probably require additional pre-fire data (such as temperature, altitudes, etc.) from sources other than the Sentinel-2 satellites. Furthermore, the project foresees leveraging the developed deep learning applications for the study of vegetation diseases in vineyards, potentially improving the effectiveness of the two methodologies within the WP6 flagship use-case.

In conclusion, the CN is a great opportunity for the Italian research, going well beyond securing cutting-edge computing resources and technologies. A fundamental aspect of the CN program is indeed the training of a new generation of computing-savvy researchers and Ph.D. graduates. These individuals will not only be qualified to address both current and emerging scientific (and societal) challenges, but will also facilitate the transfer of methods and technological solutions from academia to the private sector. Furthermore, the CN resources will be accessible to both scientific and industrial communities through flexible and standardized cloud web interfaces, seeking to form an appealing ecosystem based on strategic public-private partnerships. More broadly, the program aims to strengthen competitiveness and ensure long-term sustainability of the Italian (and European) players in the field.

Acknowledgments

This work is supported by ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by European Union – NextGenerationEU.

The authors acknowledge the fruitful discussions with colleagues of the ReD & Darkside Collaborations.

References

- [1] T. Cheng *et al.*, Phys. Rev. D 107, 016002 (2023), 10.1103/PhysRevD.107.016002.
- [2] S. Bini *et al.*, Classical and Quantum Gravity 40, 135008 (2023), 10.1088/1361-6382/acd981.
- [3] C. E. Aalseth *et al.*, Eur. Phys. J. Plus 133, 131 (2018), 10.1140/epjp/i2018-11973-4.
- [4] P. Agnes *et al.*, Eur. Phys. J. C 81, 1014 (2021), 10.1140/epjc/s10052-021-09801-6.
- [5] P. Agnes *et al.*, Phys. Rev. D 107, 063001 (2023), 10.1103/PhysRevD.107.063001.
- [6] A. Aab *et al.*, Nucl. Instrum. Meth. A 798, 172–213 (2015), 10.1016/j.nima.2015.06.058.
- [7] M. Niechciol *et al.*, PoS ICRC2023, 1488 (2023), 10.22323/1.444.1488.
- [8] J.C. Arteaga-Velázquez, PoS ICRC2023, 466 (2023), 10.22323/1.444.0466.
- [9] D. Marletta *et al.*, Technologies, 11(6), 174 (2023), 10.3390/technologies11060174.
- [10] V. Strati *et al.*, EGU24-10773 (2024), 10.5194/egusphere-egu24-10773.
- [11] R. Ghali, M. A. Akhloufi, Fire 6(5), 192 (2023), 10.3390/fire6050192.
- [12] The Pierre Auger Collaboration, *Pierre Auger Observatory Open Data*, 10.5281/zenodo.10488964.
- [13] A. Aab *et al.*, "Auger Prime: Preliminary Design Report", 2015, arXiv:1604.03637.