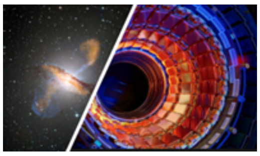

1222·2022 800 ANNI

UNIVERSITÀ DEGLI STUDI DI PADOVA

PhD course of National Interest in Technologies for
Fundamental Research in Physics and Astrophysics

Annual report

Name and surname: **Muhammad Waqas**
Cycle and a.a.: **39th**
Supervisor: **Dr.Valerio Formato**

- **Research activity carried out during the year**

Multimodal Domain-Adversarial Network for Particle 'Phosphorus (P)' Identification

Methodology and results of applying a **Multimodal Domain-Adversarial Network (MDAN)** to significantly reduce background noise for high-purity event selection of a specific particle, designated as 'P', in a physics detector experiment.

1. Motivation and Problem Statement

As part of ongoing research on AMS-02 data analysis, I successfully reproduced published results for Silicon and extended the analysis to heavier nuclei, including Phosphorus and Sulphur. To enhance signal extraction and improve the final flux measurements, the Multimodal Domain-Adversarial Network (MDAN) was implemented. This work represents a significant step toward deploying cutting-edge machine learning techniques directly within the operational data analysis chain to ensure the highest possible quality for the final flux results.

The central challenge in particle physics analysis is achieving **high signal purity** after initial event selection cuts, particularly for particles sensitive to nuclear interactions. For heavy nuclei cosmic rays, **Phosphorus ('P')** is highly sensitive to fragmentation processes, making effective background rejection an **especially crucial** requirement for accurate flux measurement. The MDAN methodology is being developed specifically for inclusion in the current **AMS-02 data analysis pipelines**.

Even with standard selection criteria applied (specifically, an L1 charge cut of 13.333 to 15.950 and an Inner Rigidity of [1,2000] GV), residual background events remain. The initial focus of this study is on rejecting the most common contaminants, specifically Z=14 (Silicon, 'Si') and Z=16 (Sulfur, 'S'). The long-term plan is to expand this generalized MDAN approach to compensate for contamination from other nuclei across the heavy cosmic ray spectrum, thereby significantly improving the overall quality of the physics sample. Currently, I am working on intensive feature engineering to make the model more robust while simultaneously ensuring it is computationally inexpensive for real-time data analysis.

2. Methodology: The Multimodal Domain-Adversarial Network

The chosen architecture, the MDAN, is a sophisticated neural network designed to achieve two critical goals simultaneously: superior classification power and robust generalization between simulated and real data.

Multimodal Feature Fusion

The network is **multimodal**, meaning it integrates complementary information from different detector systems:

1. **Tracker Information:** Processed by both a **Graph Neural Network (GNN)** (for track features like Q and hit position) and a **Convolutional Neural Network (CNN)** (for cluster ratios).
2. **Time-of-Flight (TOF) Information:** Processed by an **MLP** (for variables like Beta, Q, and clusters).
3. **Global Tracker Parameters:** Also processed by a separate **MLP** (for rigidity variables).

The features extracted by these four specialized sub-models are combined using an **attention mechanism**. This mechanism intelligently fuses the data streams, allowing the model to dynamically weight the most relevant detector components for each individual particle event.

Domain-Adversarial Training for Consistency

A major challenge is ensuring the model trained on simulated **Monte Carlo (MC)** data performs equally well on real-world **Data**. This is addressed by incorporating a **Domain-Adversarial Network (DAN)** component:

- **Signal/Background Classifier:** The primary function, which aims to maximize the distinction between the 'P' signal and background.
- **Data/MC Classifier:** A secondary function that attempts to classify whether an event came from the MC simulation or real Data.
- The training uses a **Gradient Reversal Layer** to force the feature extractor to learn **domain-invariant features**. The system is optimized so that the Data/MC classifier's accuracy hovers near 50% (a random guess). This forces the signal classifier to rely only on features that are relevant to the physics (signal vs. background), rather than relying on differences between the MC simulation and the real detector conditions.

3. Key Results and Performance

The model was trained on ~200K MC events and validated against ~160K real Data events.

Metric	Value	Interpretation
Overall AUC	0.917	The Area Under the ROC Curve is close to 1.0, indicating excellent classification performance across all possible threshold choices for the Signal/Background Classifier.
Signal Efficiency (MC)	96%	96% of true 'P' signal events are correctly identified and selected by the model, demonstrating very high purity preservation.
Background Efficiency (MC)	70%	70% of true background events (Si, S) are correctly identified as background and rejected by the model.
Data/MC Classifier Accuracy	≈50%	This value confirms the success of the Domain-Adversarial training, meaning the learned features are robust and indistinguishable between the simulation (MC) and real (Data) detector conditions.

List of attended courses and passed exams

- ❖ Introduction to Neuromorphic Computing (Material submitted)
- ❖ Big Data Modeling and Learning (Project & Technical Report is submitted)

- **List of attended conferences, workshops and schools, with mention of the presented talks 2024:**

- AMS Italia Meeting in Bologna, Italy, focused on strategic meetings regarding the Alpha Magnetic Spectrometer (AMS) Italian collaboration (November 27-29).
- NAIA Data Framework Use: Hands-on Training (Part 1) in Bologna, Italy, providing practical training on the specific data processing framework used for AMS-02 analysis (November 29).

2025:

- NAIA Data Framework Use: Hands-on Training (Part 2) in Rome, Italy, continuing the practical data framework training (February 03).
- TECH-FPA PHD Retreat 2025 at LNGS, L'Aquila, Italy, focusing on technology, funding, and career paths in particle astrophysics (February 17-21).
- Authentic Impact Workshop by Barbara Van for Team Development at LNGS, L'Aquila, Italy, focusing on maximizing authentic professional impact and collaboration (TBD).
- ISAPP School: "High Energy Cosmic Rays: From Space to Ground" in Lecce, Italy, offering intensive summer school coverage of cosmic ray physics from space-based detection to ground-based observation (June 9-20).

- AMS Nuclei Meeting at SDIAT, Jinan, China, specifically dedicated to the analysis, spectra, and interpretation of cosmic ray nuclei data from AMS-02 (September 6-12).
- 1st PhyStat School of Statistics: "Statistics in the era of ML" in the Netherlands, providing specialized training on advanced statistical methods and the application of Machine Learning (ML) in physics (November 17-21).
- TEE Shifts in POCC Data Center for AMS 02 , Geneva, Switzerland (28/03/2025 - 05/04/2025) & (11/05/2025 - 18/05/2025 & 20/09.2025 - 25/09/2025)

- **List of published papers/proceedings**

  ......

- **Thesis title (even temporary)**

Signal Enhancement and Background Modeling for AMS-02 Heavy Nuclei using Multimodal Deep Learning.

Date, 29/09/2025 **Signature:** M.waqas

Seen, the supervisor