Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Machine Learning and Deep Learning algorithms for Gaia mission data analysis

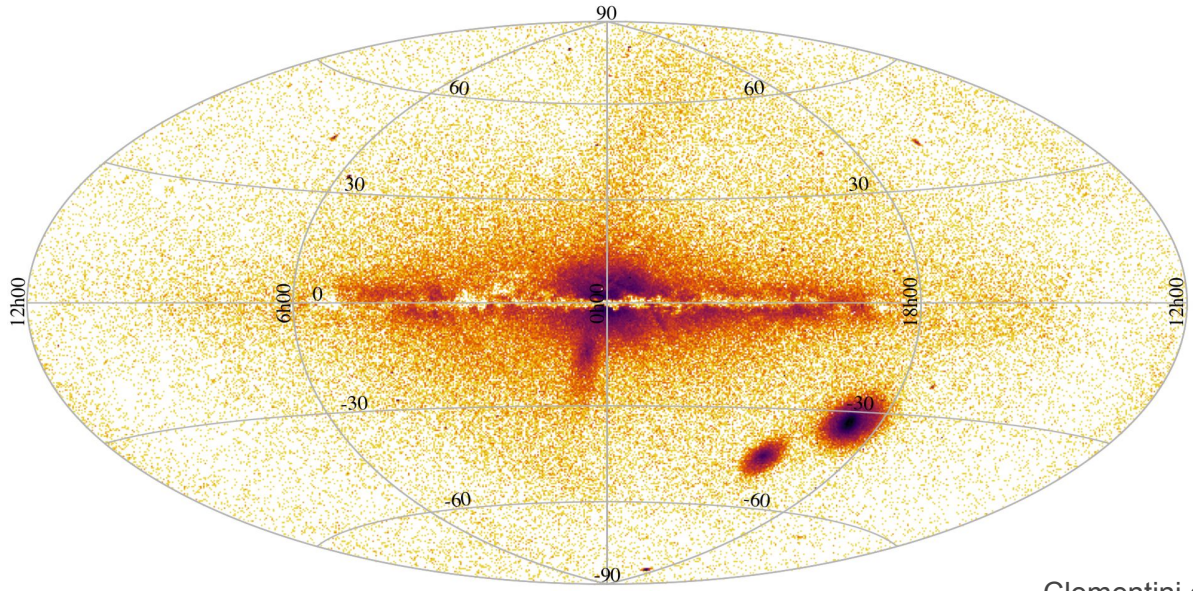
Lorenzo Monti, Tatiana Muraveva

INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna

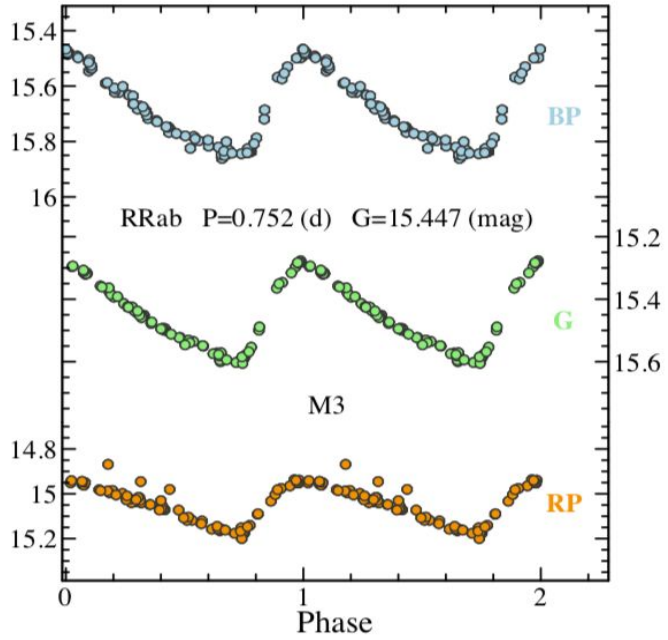
Spoke 3 Technical Workshop, Trieste October 9 / 11, 2023

Scientific Rationale

- RR Lyrae stars are periodic (Period < 1 day), pulsating, variable stars.
- There is a correlation between RRL's light curves and their metallicities ([Fe/H]).
- Gaia Data Release 3 provides a catalogue of 270 905 RRLs along with their time-series photometry.



Clementini et al. (2023)



Project Main Goal: Derive metallicities of RR Lyrae stars from their time-series photometry data using Machine Learning/Deep Learning algorithms.

Technical Objectives, Methodologies and Solutions

✔ Estimating of metallicities from the parameters of the Fourier decomposition for a limited sample (~160) of RR Lyrae stars with accurate metallicity values:

- **Collecting** the clean sample of RR Lyrae with accurate metallicity estimates.
- **Feature selection:** applying Sequential Feature Selector, Select K best to identify the most important parameters for determination of metallicity.
- **ML algorithms** (e.g. XGBoost) to measure metallicities from the Fourier parameters.
- **Bayesian** approach to fit the linear relation between Fourier parameters and metallicities.

➔ Applying all listed above methods produce a **catalogue of thousands of RR Lyrae stars** with accurate metallicity estimates.

➔ Building Neural Networks to estimate metallicity of RR Lyrae stars from time-series photometry based on the catalogue produced on the previous steps, in particular:

- Time Series Prediction with **LSTM Recurrent Neural Networks.**
- **Transformers architecture** applied to time-series instead of natural language.

Timescale, Milestones and KPIs

OCT 2022 - OCT 2023

✔ Estimating of metallicities from the parameters of the Fourier decomposition for a limited sample (~160) of RR Lyrae stars with accurate metallicity values:

- **Collecting** the clean sample of RR Lyrae with accurate metallicity estimates.
- **Feature selection:** applying Sequential Feature Selector, Select K best to identify the most important parameters for determination of metallicity.
- **ML algorithms** (e. g. XGBoost) to measure metallicities from the Fourier parameters.
- **Bayesian** approach to fit the linear relation between Fourier parameters and metallicities.

OCT 2023 - DEC 2023

Applying all listed above methods produce a **catalogue of thousands of RR Lyrae stars** with accurate metallicity estimates. Metallicity validation.

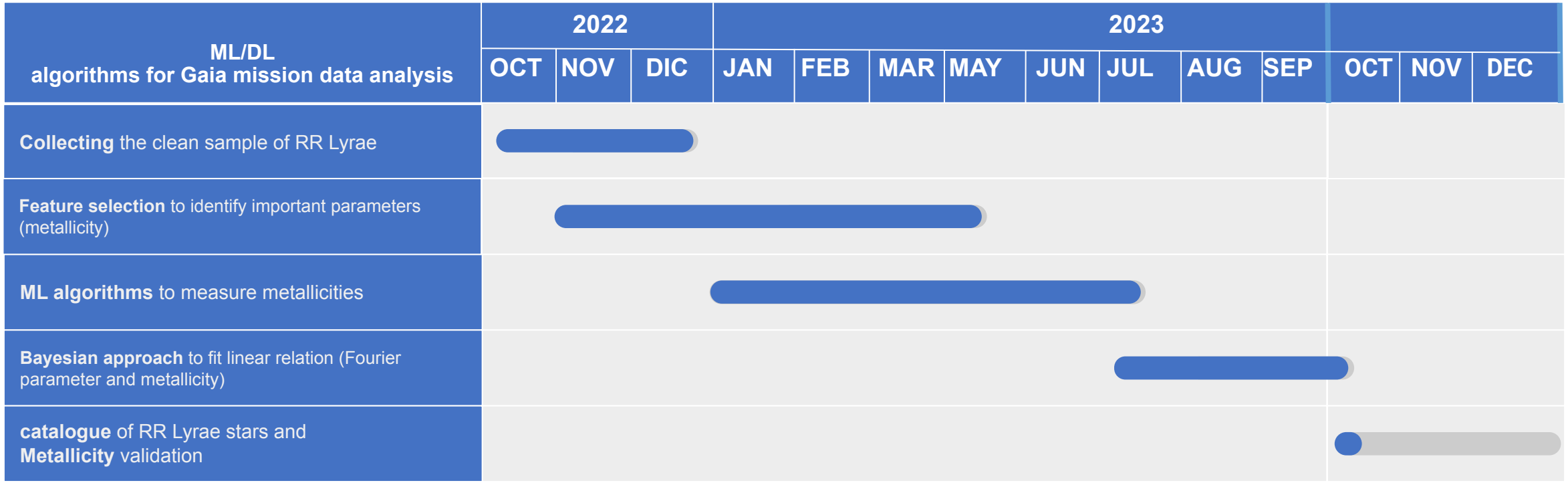
DEC 2023 – OCT 2024

Building Neural Networks such as **LSTM** and **Transformers architecture** in order to estimate metallicity of RR Lyrae stars from time-series photometry based on the catalogue produced on the step above.

Timescale, Milestones and KPIs

OCT 2023 - DEC 2023

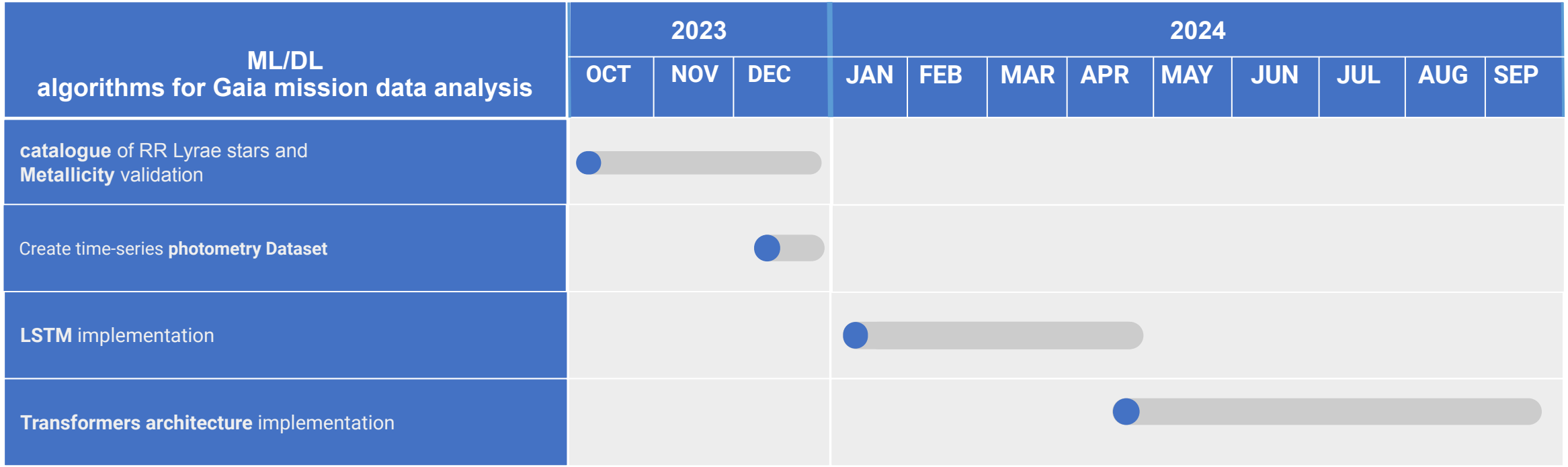
Applying all listed above methods produce a **catalogue of thousands of RR Lyrae stars** with accurate metallicity estimates. Metallicity validation.



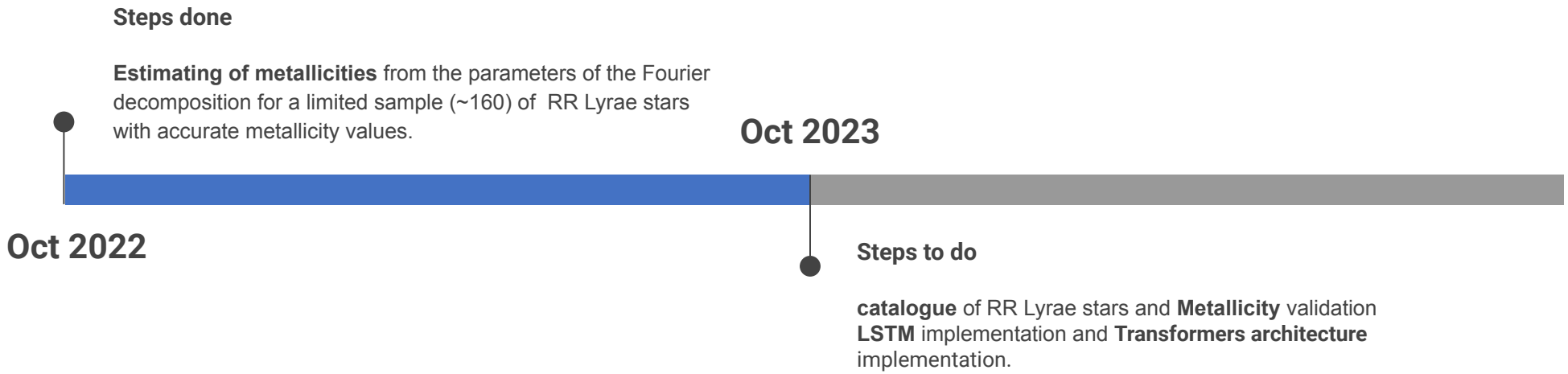
Timescale, Milestones and KPIs

DEC 2023 – OCT 2024

Building Neural Networks such as **LSTM** and **Transformers architecture** in order to estimate metallicity of RR Lyrae stars from time-series photometry based on the catalogue produced on the step (1).



Accomplished Work, Results



- Invited talk of **Muraveva T**. describing the results on IAU Symposium 376 “At the cross-roads of astrophysics and cosmology: Period–luminosity relations in the 2020s, 17-21 April 2023, Budapest, Hungary.

- Invited talk of **Muraveva T**. describing the results on the MW-Gaia 2023 meeting “Science and technology roadmap for μas studies of the Milky Way”, 18-20 July 2023, Lund, Sweden.

Next Steps and Expected Results (by next checkpoint: April 2024)

Next steps:

1) Validation of the **derived catalogue**.
2) Building a beta **LSTM model** for April (the next checkpoint).

Expected results (and KPI):

1) Paper *"Metallicities and distances of RR Lyrae stars estimated with Machine Learning algorithms"* by Muraveva et al. submitted to MNRAS.
2) Validated catalogue of RR Lyrae stars with estimated metallicities published.
3) Open source repository released with CD/CI pipeline and automatic release.

Thank you
for your attention

contact
Tatiana Muraveva tatiana.muraveva@inaf.it