PROGETTO - PROJECT	
ACRONIMO - ACRONYM	**MexicanDya**
TITOLO - TITLE	**Modern EXploration In ComputAtioNal DYnamics**
SOTTOSETTORE ERC – ERC SUBSECTOR	**PE1_20 (Applications of Math. to Sciences)**
DURATA - DURATION	**24 Months**

PROGETTO DI RICERCA - RESEARCH PROJECT

ABSTRACT (da 100 a 1500 caratteri/ from 100 to 1500 characters)

The complete simulation of the dynamics of a n-body problem in all details is often quite hard, because of the too large computational complexity of the codes producing the numerical solutions of this kind of problems. Machine learning (ML) methods have been shown to be particularly efficient in selecting nearly optimal choices for problems with a cost function depending in a complicate way by many parameters that define the system. The project MexicanDya aims to exploit such a property of ML techniques in order to either introduce properly a simplified model or a good set of parameters, in order to drastically reduce the computational cost of the simulations. This approach is expected to allow the study of the dynamics of interesting kinds of n-body problems for a so extended timespan, that is otherwise unreachable. As a ultimate goal, MexicanDya aims to highlight interesting phenomena of the long-term dynamics in the framework of the so cleaverly designed mathematical models.
Since MexicanDya considers multi-disciplinarity as a target to be strongly promoted, the previously described approach will be applied to problems of interest for different natural sciences: Astronomy (in the context of an approach that is usual in the field of Mathematical Physics), Chemistry and Biology.

STATO DELL'ARTE - STATE OF THE ART
(da 100 a 3000 caratteri/ from 100 to 3000 characters)

Since MexicanDya aims to apply Machine Learning (ML) methods in the context of different problems of computational dynamics, it is worth to subdivide the present state-of-the-art section in some paragraphs in order to briefly recall the framework for a few of these different topics.
Neural networks have to be considered as the core tool of artificial intelligence and are used to efficiently implement ML techniques. Software packages originally designed to use ML for image recognition were adapted to generate generate particles trajectories in turbulent systems, by avoiding (heavy) numerical simulations and in such a way to well reproduce key statistical features [Li+].
Nowadays, nearly 6000 confirmed exoplanets are listed in the NASA's Archive; they were

detected by using different detection method, but none of them is able to provide information about all the orbital elements. For instance, the radial velocity (RV) technique is completely blind to the (mutual) inclinations, that strongly affect the dynamics of the extrasolar planetary systems. The missing information can be retrieved by applying a so called "reverse approach": among all the possible initial data, we select ranges of orbital elements values such that the dynamical evolution of the exoplanetary system is stable. In [LCSV] the criterion of the "minimal area" was introduced in order to recognize initial conditions generating very robust orbits. In the case of a system hosting n exoplanets, a careful search of the minimal area would require to investigate 3nD grids with very narrow meshes. The criterion is expected to work because a rough search of the minimal area allowed to locate initial conditions corresponding to motions that were proved to be stable according to KAM theory [CLSV1,ML2].
In general, for the n-body problem and, in particular, for what concerns Molecular Dynamics (MD), the optimisation of the simulations is a critical task in computational research, directly impacting the efficiency and accuracy of studies in fields like chemistry, biophysics, etc. GROMACS, a leader in MD simulation software, offers extensive parallelisation capabilities and fine-tuning options. However, existing performance enhancement practices often rely on heuristic methods or One-Factor-At-a-Time (OFAT) approach, which limit the exploration of complex parameter interactions and may lead to suboptimal configurations [GP]. The systematic optimisation of GROMACS performance remains underexplored despite advancements in computational techniques and HPC infrastructure. Previous studies showed the benefits of parameter optimisation in related domains, yet a structured application of these methods to large-scale MD simulations is lacking.
[Li+] T. Li et al.: "Synthetic Lagrangian turbulence by generative diffusion models", Nature Machine Intelligence, 6 (2024).
[GP] C. Gruber, J. Pleiss: "Systematic benchmarking of large MD simulations employing GROMACS ...", J. Comput. Chem., 32 (2011).

DESCRIZIONE DEL PROGETTO - PROJECT DESCRIPTION
(da 1000 a 5000 caratteri / from 1000 to 5000 characters)

MexicanDya aims to study a class of n-body problems with the same "two stages" approach, where Machine Learning (ML) methods are heavily used to conveniently design the modelization into the first (preliminar) step. All the problems that are considered as targets for the project share a common basic structure: they are represented by a system of ordinary differential equations (often of Hamiltonian type) which determine the motion of the particles making part of the system under study. Moreover, the dynamical evolution of these systems can be conveniently approximated (with a procedure that can be either analytical or mainly numerical) by a simplified model depending on a set of parameters. Here, ML comes into play, because it allows a fine tuning of the choice on the values of those parameters in such a way to greatly increase the reliability of the model and/or the performances of the numerical simulations.
The application of this "two stages" strategy to the study of the orbital dynamics of the extrasolar planetary systems is a paradigmatical example for explaining our approach. We plan to train a ML software package to detect initial conditions corresponding to minimal areas and so to robust orbital configurations, that are expected to be the most probable ones in particular for the subsystem formed by the biggest exoplanets. This is because the motion of small bodies

would be perturbed so much by major planets in a marginally stable situation that they would be ejected by the system in not so long times. Often there is a natural hierarchy with a subsystem (formed by the star and a few other large bodies) on top of the list. After having determined the most robust orbital configuration for the main subsystem the procedure can be restarted by scrolling the hierarchy from top to down, so as to include smaller and smaller exoplanets in nested models that are restricted because the lastly introduced exoplanet is assumed to not exert any significant gravitational attraction on the others. Remarkable results were already obtained by implementing this hierarchical scheme with a semi-analytical technique [ML1,ML2], which is so computationally demanding that makes the introduction of faster methods essentially unavoidable to widely explore the orbital dynamics of exoplanets.
Nonlinear Hamiltonian systems admit regular and chaotic solutions depending on the choice of system parameters and initial conditions. The performance of numerical methods to approximate the solutions strongly depends on the type of solution. Simulations usually require an a priori phase of testing to fine-tune the method parameters depending on the chaoticity of the region of interest. We plan to train neural networks during this initial simulation phase for various dynamical problems, with the aim to amend classical methods by ML techniques predicting the optimal choice of the method parameters. A ML-based approach can also distinguish between regular and chaotic regions on the fly during numerical simulations.
MexicanDya aims to study simple dynamical models of interest in astronomy, biology, and chemistry. The compilation of standardized tests through several topics is expected to produce a reference dataset to be used for the training of neural networks for various applications.
In Molecular Dynamics (MD) simulations, the performance optimisation requires to explore the multidimensional computational parameters space systematically. MexicanDya adopts a structured Design of Experiment (DoE) methodology to optimise computational settings across diverse systems and configurations. Our methodology involves testing combinations of key GROMACS parameters across molecular systems ranging in size from 5,000 to 1 million atoms. The use of statistical analyses will identify critical parameters and their interactions, guiding optimal configurations for diverse simulation environments. A main goal of this study is to develop a ML-based dataset capable of predicting optimal parameters, thereby facilitating scalable and efficient simulations. Automating the performance tuning process will not only streamline simulation workflows but also establish robust best practices applicable across a broad spectrum of computational explorations.
For what concerns biological applications of MD, the large number of particles and degrees of freedom, combined with the short time scale of the relevant atomic motions, results in a limited time frame that can be sampled by this technique. This is a great limitation for the study of these dynamical systems in the timespan corresponding to many crucial processes involving biomolecules: protein conformational changes, protein-complex formation, etc. Several approaches are now being proposed to blend ML approaches with classical force fields, leveraging on the acquired data on MD runs that can be used to train a ML system to recapitulate the dynamics of the macromolecules in a computationally simplified scheme.

OBIETTIVI E RISULTATI ATTESI - OBJECTIVES AND EXPECTED RESULTS
(da 100 a 3000 caratteri / from 100 to 3000 characters)

As already discussed in the previous section, the MexicanDya project is based on a "two stages" approach, where Machine Learning (ML) methods are used to carefully design the modelization into the first step. Moreover, the second stage is devoted to the study of the long-term dynamics (and other puzzling features) of the models previously derived. In the previous section, we also described possible applications that are of interest in Mathematics, Physics/Astronomy, Chemistry, Biology. We expect to produce results in all these fields. This makes evident that MexicanDya is particularly ambitious, because of the impressively wide spectrum of the covered topics. Nevertheless, our project is feasible and it is worth to spend some words to support this (challenging) statement.
The background of the members of the research group covers nearly all the disciplines of faculty of science, because they got Master Degrees / PhD titles in Mathematics, Physics, Astronomy, Astrophysics, Chemistry, Bioinformatics. We also emphasize that in the MexicanDya team there are all the competences to deal with the High Performance Computing (HPC) techniques that must be implemented in order to carry on this research project. All the team members are experienced programmers who usually design applications to scientific computing. Among us there are developers using software packages like GROMACS, PySINDy, Mathematica, Tensorflow, etc. whose knowledge is essential for the research project. There are also substantial competences about long-term numerical integrations of ordinary differential equations. Moreover, in the last few years, some of us actively contributed to the configuration of the hardware (and the software) hosted in the computer center of the Dept. of Mathematics, that was designed to let have access to the GPU cards, which are so important to run efficiently ML codes. We also stress that we plan to not limit to use these software packages as blackboxes. In particular, since the nodes of the neural networks are working as automata (in a mathematical sense), we plan to better understand the theoretical framework of ML methods, exploiting the competences in computer science of some of us.
Last but not least, we emphasize that the results we expect are not limited to the publication of (hopefully) interesting scientific papers. Nearly all the members of MexicanDya actively participate to the meetings of the "Colloqui interdipartimentali sui nuovi metodi computazionali". In our opinion, MexicanDya will significantly enhance this kind of collaboration, that is devoted to the development of HPC methods in a transversal way with respect to the Departments of the Faculty of Science of the Univ. of Rome "Tor Vergata".

PAROLE CHIAVE - KEY WORDS

Mathematical Models
N-Body Problems
Computational Dynamics
Stability and Chaoticity in Hamiltonian Systems
Numerical Simulations
High Performance Computing (HPC)
Machine Learning (ML)
Multi-disciplinarity
Applications of Mathematics to Sciences