europass

Melika Keshavarzmirzamohammadi

Date of birth: 20/09/1995 | **Nationality:** Iranian | **Phone number:** (+39) 3343022241 (Mobile) | **Email address:** Melikasupernova@gmail.com | **Website:** https://github.com/Melikakmm |

Address: Via Guido Puchetti, 2, 35129, Padova, Italy (Home)

ABOUT ME

After receiving my Master's degree from the University of Padova in October 2023, I was awarded a pre-doctoral research position at the same university in computational biophysics. My research focuses on implementing cutting-edge machine learning methods on disordered proteins and their structures. In addition to my academic background in biology, physics, and mathematics, I have undergone extensive training in various computational areas, such as machine learning, deep learning, data analysis and statistics, data management, and distributed computing.

DIGITAL SKILLS

Python | c++ | Cython | R programming language | SQL | MySQL | Docker | Pytorch | Keras | TensorFlow | scikit-learn | Git / Gitlab / Github | Distributed computing, Dask | Linux | Graph Neural Networks (GNN) | expert user of pymol | LinkP | Molecular Simulation | TRANSFORMERS | biopython | Nature Language processing | Knowledge of Language Models and Transformers in NLP | Hugging Face Datasets | Cloud-computing | Visual Studio

LANGUAGE SKILLS

Mother tongue(s): **PERSIAN**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C1	C1	C1	C1
GERMAN	A1	A1	A1	A1	A1
ITALIAN	A1	A1	A1	A1	A1

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user

EDUCATION AND TRAINING

09/2023 – CURRENT Padova, Italy
PRE-DOCTORAL RESEARCHER Padova university

10/2021 – 09/2023 Padova, Italy
MSC IN PHYSICS OF DATA University of Padova

Website https://www.unipd.it/en/

09/2014 – 06/2018 Tehran, Iran
BSC IN PHYSICS University of Alzahra

Website https://en.alzahra.ac.ir/

2017 Zanjan, Iran
PHYSICS WINTER SCHOOL Institute for Advanced Studies in Sciences

Website https://iasbs.ac.ir/

WORK EXPERIENCE

10/2023 – CURRENT Padova, Italy
SCIENTIFIC RESEARCHER (COMPUTATIONAL BIOPHYSICIST) THE UNIVERSITY OF PADOVA (PROF. AMOS MARITAN)

The research includes two important projects:

1. Despite breakthroughs such as AlphaFold2 addressing the challenge of estimating protein structures for complexes to some extent, the structural estimation for disordered proteins and the dynamics of these proteins remains an even more significant challenge. In contributing to this endeavour, I have developed a Python module capable of extracting essential information to construct our dataset. Additionally, I have created a set of residue dictionaries used to embed residues with their intrinsic properties into arrays suitable for inputting deep learning methods. Furthermore, I have conducted extensive data analysis on pair residues and their frustration index in specific protein complexes. Currently, we are engineering a cutting-edge attention-based neural network specifically tailored to tackle the structural challenges posed by disordered proteins. This neural network utilizes attention mechanisms to provide context to input residue chains, attributing significance to the arrangement of residues that form a particular structure.
2. The second project examines droplet size distribution below the transition concentration for protein phase separation, which classical nucleation theory cannot readily explain. We analyzed the droplet size distributions as a function of protein concentration and found they follow a scale-invariant log-normal pattern, with the average size increasing as the concentration approaches the transition point. Collaborating with Cambridge University, our team at the University of Padova conducted experiments on proteins such as FUS, alpha-synuclein, and Arc to confirm this log-normal distribution and derive universal coefficients.

The GitHub repositories associated with these projects are presently set to private access. This confidentiality has been implemented at the team's request, primarily in anticipation of forthcoming publications.

10/2023 Padova, Italy
MY THESIS PROJECT: TESTING DEEP LEARNING AND MACHINE LEARNING METHODS TO ESTIMATE THE METALLICITY OF RR LYRAE STARS FROM THE FOURIER TRANSFORM PARAMETERS OF THEIR LIGHT CURVES (SUPERVISED BY DR GIULIANO IORIO)

RR Lyrae stars are invaluable as chemical tracers due to their well-established link between heavy-element abundance and light curve features. However, achieving precise and consistent calibration across diverse photometric wavelengths has posed a persistent challenge. Numerous pioneering studies have proposed cutting-edge techniques to address this issue. In this project, we employ a combination of machine learning and three deep learning approaches. Our machine learning approach explores the feasibility of predicting metallicity from the Fourier Transform parameters of RRL light curves, while the three deep learning methods aim to perform regression from one photometric band to another. In essence, we first employ an existing metallicity prediction method on extensive photometric datasets, followed by the training of three Recurrent Neural Network (RNNs) to regress [Fe/H] values based on light curves in alternate wavelength bands. These three RNNs include the Long Short-Term Memory RNN, the Gated Recurrent Unit (a faster and simpler version of LSTM), and a transfer-learned LSTM that enhances the capabilities of the initial LSTM for the metallicity prediction of new Gaia data release.

Link https://github.com/Melikakmm/Master_Thesis/tree/main

Padova, Italy
GAUSSIAN PROCESS REGRESSION FOR FITTING INTERATOMIC POTENTIALS (SUPERVISED BY MONIKA FUXREITER)

Gaussian Process Regression is one of the most significant regression techniques used not only to capture intricate details within data but also to effectively filter out unwanted noise. This method has demonstrated its flexibility and utility across various applications in physics, ranging from astronomy to molecular simulations. The strength of this model lies in its adaptability, as it can be fine-tuned to match the data's

behaviour by adjusting hyperparameters and selecting appropriate kernels. This project comprises two primary objectives. First, we utilize GPR to estimate the quantum mechanical energy of water molecules within the Born-Oppenheimer approximation. Second, we manually design a GPR model tailored to 400-atom clusters composed of argon, ranging in size from 12 to 19 atoms. In this scenario, the GPR model is implemented to predict the potential energy of the clusters by simplifying the atomic interactions inside each cluster to pairwise interactions. Each implementation is optimized to identify the most suitable hyperparameters for the model with respect to the data.

Link https://github.com/Melikakmm/GPR_fitting_interactive_potential/blob/main/fitpot.ipynb

CURRENT

TRANSLATING ENGLISH TO TWO LANGUAGES USING ATTENTION BASED MODELS (NLP) (TRANSFORMERS) SELF-DIRECTED PROJECT

I have successfully reproduced the groundbreaking transformer model from the paper "Attention is All You Need" to translate English into two languages: Italian and Persian. This is just the beginning of an exciting journey!

1. First, I'm eager to explore whether alphabetically similar languages, like English and Italian, are easier to learn compared to vastly different alphabets, such as English to Persian.
2. Next, I plan to develop an algorithm that can handle typos and wrong grammar. This will make translations even more accurate and reliable.
3. My third objective is to create a new natural language model capable of generating British humour based on a given context or topic after the first and second objectives (DeepJoke project).

This project aspires to push the boundaries of my understanding of NLP models, embracing groundbreaking approaches, and gaining priceless experience along this exhilarating journey.

Link https://github.com/Melikakmm/NLP_Translation

CURRENT Padova, Italy

GRAPH NEURAL NETWORK IMPLEMENTATION ON THE STRUCTURE OF PROTEIN COMPLEXES TO PREDICT FRUSTRATION INDEX (SUPERVISED BY PROF.MONIKA FUXREITER)

Unraveling the intricate nature of protein-protein interactions and their remarkable folding dynamics poses a formidable scientific challenge. Investigating the precise mechanisms underlying the binding and folding within protein complexes requires a detailed understanding of the connections between different regions. This study delves into the multifaceted aspects of these connections to answer three types of questions. Firstly, it investigates how the identification of different amino acid types and the precise location of residues influence the connection between them. Secondly, it explores the impact of characterizing neighboring bounds on the overall stability and functionality of the connections. Lastly, it explores the energetic landscape that governs these interactions. To tackle these challenging questions, we employed Graph Neural Networks (GNNs) for several reasons. Firstly, amino acid types, residue numbers, and chain types can be encoded as features of each node. Secondly, it has the versatility of considering the impact of other nodes on a single node, and finally, this method provides the flexibility to train and predict the edges properties instead of nodes, which is very useful for this field of study, since its main focus is about the energetic properties of the connections between the nodes(or the residues). In this study, these energetic properties are described with a measure called frustration, which hinges on a ratio of the energy difference between the native structure from alternatives to the magnitude of the fluctuations of the decoy energies. In this project, we use a GNN technique in order to predict the frustration index (edges) between two residues (nodes) with highlighted properties(features). After presenting this primary project idea, my supervisor Prof. Monika Fuxreiter offered me a research position for implementing AI-based techniques to estimate disordered protein structures.

The project remains in progress. Currently, I am dedicating time to improving this project. Since the original dataset did not include the frustration index for all contacts, we are working on producing a more inclusive and meaningful dataset for this task.

Padova, Italy

CNN FOR AUDIO CLASSIFICATION PROJECT (SUPERVISED BY DR.JACOPO PEGORARO)

In this project, we tackled an audio classification problem using Convolutional Neural Networks (CNN) in Pytorch. We approached the problem as an image classification task by utilizing mel-spectrograms of the music. The project had several objectives. First, we created a custom data class, which processed and converted the raw data into mel-spectrograms. Second, we experimented with various well-known NN architectures, as well as our own architectures, to compare their performance. Third, we discovered that all

the pre-trained weights already available in PyTorch were trained on image classification problems rather than music. Therefore, we attempted to create our own pre-trained weights on a similar dataset (GTZAN) to use for our task. Finally, our last goal was to develop a Sound Fusion architecture that could learn from different audio representations.

Link https://github.com/Melikakmm/CNN-for-sound-classification

Padova/London

TESTING MACHINE LEARNING METHODS FOR VOLATILITY-CARRY-TRADING-STRATEGY PROJECT (REMOTELY SUPERVISED BY XSOR CAPITAL TEAM, LONDON)

The main goal of this project was to develop a volatility carry trading strategy that can be used in the foreign exchange (forex) market. The project aims to create a trading algorithm that is capable of predicting market volatility and generate profitable trades based on this prediction. To achieve this goal, the project involves several steps including data preprocessing, feature engineering, model training, and backtesting. The project also aims to compare the performance of different machine learning models and select the best model for the trading strategy. Additionally, the project seeks to analyze the risks and limitations of the trading strategy and provide recommendations for future improvements.

Link https://github.com/Melikakmm/Volatility-carry-trading-strategy

Padova, Italy

PYTHON/CYTHON WRAPPER PROJECT FOR SEVN (SUPERVISED BY DR.GIULIANO IORIO AND PROF. MICHELA MAPELLI FROM UNIVERSITY OF PADOVA)

SEVN, also known as Stellar EVolution N-body, is a high-speed population synthesis program created using C++. Its main objective is to simulate the evolution of binary systems, where the characteristics of the entire system are significantly influenced by the stars within it and their interactions. This project seeks to combine the speed and efficiency of C++ with the user-friendly nature of Python. The ultimate goal is to create a wrapper for the SEVN software that enables users to interact with it seamlessly, without having to grapple with the complexities of the C++ language. By leveraging the strengths of both languages, users can benefit from the impressive performance offered by C++, while enjoying the ease of use and accessibility provided by Python

Link https://github.com/Melikakmm/SEVN_PYTHON_WRAPPER

Padova, Italy

LOW-MASS-X-RAY-BINARIES PROJECT (SUPERVISED BY DR.GIULIANO IORIO AND PROF. MICHELA MAPELLI)

In this research project, we focused on studying Low-Mass-X-Ray Binary Systems (LMXBs) by utilizing data extracted from SEVN, a population synthesis code. However, the initial dataset was very broad, existing in multiple files, and encompassing every type of binary system. In order to hone in on the LMXBs, a filter was developed in Python, which took into account the theoretical and empirical constraints applicable to such systems. Once we had obtained the LMXBs data, we conducted multiple analyses on the properties of the LMXBs in order to understand more about the evolution and initial conditions of such systems. Furthermore, I have devised an algorithm capable of automating the organization, cleansing, and analysis of data extracted from SEVN.

Link https://github.com/Melikakmm/Low-Mass-X-ray-Binaries

Padova, Italy

INFORMATION THEORY PROJECT (PROF. MICHELE ALLEGRA)

Exploring the information flow in a **pollen** and **pollinator** network using Infomap for the Ashu forest pollination network.

Link https://github.com/jjackson1994/Information_Theory

02/2018 – 05/2021 Tehran, Iran

OLYMPIAD PHYSICS TEACHER FARZANEGAN HIGHSCHOOL

In February 2018, working together with two collaborators, I formed a group to prepare students for the Physics Olympiad competition in Iran. We were successful and found employment in two schools

immediately (Farzanegan and AE high schools). The material covered in these courses was equivalent to approximately half of a bachelor's degree in physics. Given the students' need for improvement in mathematics to understand many disciplines in physics, the courses proved to be highly challenging for both students and instructors, requiring a significant level of intensity in teaching and learning. In addition to teaching, it was my role to assign tutorial responsibilities and coordinate meetings with various schools and institutes. Farzanegan High School is part of The National Organization for Development of Exceptional Talents (NODET) which is funded by funded by the Ministry of Education. Many famous figures, like the Fields Medal winner Maryam Mirzakhani, are noted as alumni. I continued working in this position until I moved to Italy for my master's degree.

02/2018 – 06/2021 Tehran, Iran

OLYMPIAD PHYSICS TEACHER AE HIGHSCHOOL

- teaching physics and mathematics to Olympiad Physics students
- assigning tutorial responsibilities.
- coordinating meetings with various schools and institutes.