



BIOTECHNO 2023

The Fifteenth International Conference on Bioinformatics, Biocomputational
Systems and Biotechnologies

ISBN: 978-1-68558-058-2

March 13th - 17th, 2023

Barcelona, Spain

BIOTECHNO 2023 Editors

Lorena Parra, Universitat Politecnica de Valencia, Spain

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany

BIOTECHNO 2023

Foreword

The Fifteenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies (BIOTECHNO 2023), held between March 13 – 17, 2023, covered these three main areas: bioinformatics, biomedical technologies, and biocomputing.

Bioinformatics deals with the system-level study of complex interactions in biosystems providing a quantitative systemic approach to understand them and appropriate tool support and concepts to model them. Understanding and modeling biosystems requires simulation of biological behaviors and functions. Bioinformatics itself constitutes a vast area of research and specialization, as many classical domains such as databases, modeling, and regular expressions are used to represent, store, retrieve and process a huge volume of knowledge. There are challenging aspects concerning biocomputation technologies, bioinformatics mechanisms dealing with chemoinformatics, bioimaging, and neuroinformatics.

Biotechnology is defined as the industrial use of living organisms or biological techniques developed through basic research. Bio-oriented technologies became very popular in various research topics and industrial market segments. Current human mechanisms seem to offer significant ways for improving theories, algorithms, technologies, products and systems. The focus is driven by fundamentals in approaching and applying biotechnologies in terms of engineering methods, special electronics, and special materials and systems. Borrowing simplicity and performance from the real life, biodevices cover a large spectrum of areas, from sensors, chips, and biometry to computing. One of the chief domains is represented by the biomedical biotechnologies, from instrumentation to monitoring, from simple sensors to integrated systems, including image processing and visualization systems. As the state-of-the-art in all the domains enumerated in the conference topics evolve with high velocity, new biotechnologies and biosystems become available. Their rapid integration in the real life becomes a challenge.

Brain-computing, biocomputing, and computation biology and microbiology represent advanced methodologies and mechanisms in approaching and understanding the challenging behavior of life mechanisms. Using bio-ontologies, biosemantics and special processing concepts, progress was achieved in dealing with genomics, biopharmaceutical and molecular intelligence, in the biology and microbiology domains. The area brings a rich spectrum of informatics paradigms, such as epidemic models, pattern classification, graph theory, or stochastic models, to support special biocomputing applications in biomedical, genetics, molecular and cellular biology and microbiology. While progress is achieved with a high speed, challenges must be overcome for large-scale bio-subsystems, special genomics cases, bio-nanotechnologies, drugs, or microbial propagation and immunity.

We take here the opportunity to warmly thank all the members of the BIOTECHNO 2023 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to BIOTECHNO 2023.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the BIOTECHNO 2023 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that BIOTECHNO 2023 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of bioinformatics, biocomputational systems and biotechnologies.

We are convinced that the participants found the event useful and communications very open. We also hope that Barcelona provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city

BIOTECHNO 2023 Chairs:

BIOTECHNO 2023 Steering Committee

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany

Hesham H. Ali, University of Nebraska at Omaha, USA

BIOTECHNO 2023 Publicity Chairs

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

BIOTECHNO 2023

Committee

BIOTECHNO 2023 Steering Committee

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Hesham H. Ali, University of Nebraska at Omaha, USA

BIOTECHNO 2023 Publicity Chairs

Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain
José Miguel Jiménez, Universitat Politecnica de Valencia, Spain

BIOTECHNO 2023 Technical Program Committee

Behrooz Abbaszadeh, University of Ottawa, Canada
Antonino Abbruzzo, University of Palermo, Italy
A M Abirami, Thiagarajar College of Engineering, Madurai, India
Don Adjero, West Virginia University, USA
Aftab Ahmad, City University of New York, USA
Jens Allmer, Hochschule Ruhr West - University of Applied Sciences, Germany
Joel P. Arrais, University of Coimbra, Portugal
Simone Avesani, Università di Verona, Italy
Yoseph Bar-Cohen, Electroactive Technologies / NDEAA Lab - Jet Propulsion Laboratory (JPL), USA
Kais Belwafi, King Saud University, Saudi Arabia
Boubaker Ben Ali, University of Bordeaux, France / University of Manouba, Tunisia
Razvan Bocu, Transilvania University of Brasov, Romania
Vincenzo Bonnici, Università di Parma, Italy
Paolo Cazzaniga, University of Bergamo, Italy
Matthias Chung, Virginia Tech, USA
Peter Clote, Boston College, USA
Matteo Comin, University of Padova, Italy
Giovanni Cugliari, University of Turin, Italy
Adithi Deborah Chakravarthy, University of Nebraska at Omaha, USA
Maria Evelina Fantacci, University of Pisa, Italy
Abdolhossein Fathi, Razi University, Iran
Rosalba Giugno, University of Verona, Italy
Henry Griffith, University of Texas at San Antonio, USA
Md Shafaeat Hossain, Southern Connecticut State University, USA
Chih-Cheng Hung, Kennesaw State University, USA
Asier Ibeas, Universitat Autònoma de Barcelona, Spain
Dimitrios G. Katehakis, Center for eHealth Applications and Services FORTH | Institute of Computer Science, Greece
Jan Kubicek, VSB - Technical University of Ostrava, Czech Republic
Antonio LaTorre, Universidad Politécnica de Madrid, Spain
Chen Li, Monash University, Australia

Yiheng Liang, Bridgewater State University, USA
Tatjana Lončar-Turukalo, University of Novi Sad, Serbia
Dulani Meedeniya, University of Moratuwa, Sri Lanka
Bud Mishra, New York University (NYU), USA
Chilukuri K. Mohan, Syracuse University, USA
Constantin Paleologu, University Politehnica of Bucharest, Romania
Eleftheria Polychronidou, Information Technologies Institute - Centre for Research and Technology Hellas (ITI/CERTH), Greece
Vincent Rodin, University of Brest, France
Ulrich Rueckert, Bielefeld University, Germany
Thomas Schmid, Universität Leipzig, Germany
Andrew Schumann, University of Information Technology and Management in Rzeszow, Poland
Shima Shafiee, Razi University, Iran
Christine Sinoquet, University of Nantes, France
Elmira Amiri Souri, King's College London, UK
Andrea Tangherloni, University of Bergamo, Italy
Angelika Thalmayer, Friedrich-Alexander University Erlangen-Nürnberg, Germany
Rama Krishna Thelagathoti, University of Nebraska at Omaha, USA
Manuel Tognon, Università di Verona, Italy
Asma Touil, Université de Sousse, Tunisia / IMT Atlantique, France
Markos G. Tsipouras, University of Western Macedonia, Greece
Sophia Tsoka, King's College London, UK
Magda Tsolaki, Aristotle University of Thessaloniki, Greece
Arban Uka, Epoka University, Albania
T. Venkatesan, Sanskrithi School of Business, Puttaparthi, India
Eva Viesi, Università di Verona, Italy
Panagiotis Vlamos, Ionian University, Greece
Can Wang, Leiden University, Netherlands
Ehsan Yaghoubi, University of Beira Interior, Covilhã, Portugal
Ahmad Anwar Zainuddin, International Islamic University Malaysia, Malaysia
Elena Zaitseva, University of Zilina, Slovakia
Erliang Zeng, University of Iowa, USA
Haowen Zhang, Georgia Institute of Technology, USA
Qingxue Zhang, Purdue University, USA
Qiang Zhu, The University of Michigan - Dearborn, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission to reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

A Microservice-oriented AI Automation Framework for Supporting Single-cell Downstream Analysis <i>Hong Qing Yu, Ali Kermanizadeh, Sam O'Neill, and Oyetola Florence Idowu</i>	1
Machine Learning in the Identification of Key Residues of Variants and Polymorphisms in the Interaction of ACE2 Proteins with Spike of SARS-CoV2 <i>Ana Carolina Damasceno Sanches, Ana Luisa Rodrigues de Avila, Levy Bueno Alves, and Silvana Giuliatti</i>	7
Analyzing Switch Regions of Human Rab7a and Rab10 by Molecular Dynamics Simulations <i>Levy Bueno Alves, Sarah Sandy Sun, and Silvana Giuliatti</i>	12
Improving Recovery Engagement for Patients with Substance Use Disorder in the Emergency Department <i>Kory London and Les Sztandera</i>	16
Examining the Relationship between COVID-19 Mobility and Eviction Rates in Philadelphia <i>Regina Ruane and Les Sztandera</i>	21
Altering the Behaviour of the <i>Catostylus mosaicus</i> Jellyfish using Electromagnetic Fields <i>Alberto Ivars, Francisco Javier Diaz, Lorena Parra, Eva S. Fonfria, Cesar Bordehore, Sandra Sendra, and Jaime Lloret</i>	25

A Microservice-oriented AI Automation Framework for Supporting Single-cell Downstream Analysis

Hong Qing Yu
School of Computer Science and Engineering
University of Derby
Derby, United Kingdom
Email: h.yu@derby.ac.uk

Ali Kermanizadeh
School of Human Sciences
University of Derby
Derby, United Kingdom
Email: a.kermanizadeh@derby.ac.uk

Sam O'Neill
School of Computer Science and Engineering
University of Derby
Derby, United Kingdom
Email: s.oneill@derby.ac.uk

Oyetola Florence Idowu
School of Computer Science and Engineering
University of Derby
Derby, United Kingdom
Email: o.idowu4@unimail.derby.ac.uk

Abstract—Single-cell analysis has real potential for reshaping the future of biomedical research allowing for a better understanding of the natural properties of both healthy and diseased tissues that, in turn, allow for better opportunities for overcoming current challenges in drug discovery, diagnostics and prognostics. A large body of research in this field produces large quantities of data. Merging with fast-developed Machine learning (ML) and Artificial Intelligent (AI) algorithms would allow single-cell analysis research to be conducted more efficiently and accurately than currently possible. Therefore, there has been a surge of ML and AI developments for the whole life cycle of the downstream single-cell analysis process. However, there is a limitation to reusing, exchanging, sharing, applying the most advanced technologies, and automating the experimental environments and outcomes in cross-disciplinary collaborative research. This paper presents an automation framework to address these limitations and shows how AI and ML research can contribute to biomedical automation and control. Moreover, the real-world case will be evaluated to demonstrate the prototype implementation at the end of the paper.

Index Terms—Single-cell analysis, RNA-Seq, Machine Learning, AI Automation, Downstream Analysis, Knowledge Graph

I. INTRODUCTION

Since the turn of the century, bio-informatics research underwent a technological revolution allowing the generation of single-cell genomics data. One of the most common and fast-developed techniques is single-cell Ribonucleic acid sequencing (RNA-seq) which represents quantification and profiling of the changing gene expressions in single cells and how they differ across thousands of cells within a heterogeneous sample [1]. Single-cell genomics offers a unique opportunity to allow joint multidisciplinary research across multiple types of datasets [2]. The generation of large datasets has its own challenges that include but are not limited to sharing and reusing computational methods, algorithms, pipelines, and other resources. In addition, the automatic creation of a new analysis process composing existing research outcomes to

solve a given task is crucial to fast-track the research output, and dissemination [3], [4]. Our research aims to provide an automatic AI framework that could address the challenges in single-cell downstream data analysis including data annotation, data quality control, data normalisation, data dimensional reduction, and data analysis [5]. In data analysis, cutting-edge methods such as Machine learning algorithms can be applied for more efficient and meaningful classification, clustering, segmentation, and prediction [6]. With the development of Deep Neural Networks (DNN), models have been developed for single-cell analysis which creates further technology burdens to reuse and deploy the solutions by non-programming-focused researchers [7].

The paper has two main contributions to make the downstream data analysis more shareable and reusable and automatically provide new solutions creatively:

- 1) The developed analysis methods can be shared and reused as semantic microservices registered in the framework with annotations. Therefore, the microservices can be dynamically allocated and composed later to perform new and suitable tasks.
- 2) A semantic knowledge representation and learning framework is developed. The framework can learn the context knowledge of the tasks and dynamically search, compose, and run the registered microservices to complete future (similar) requested tasks.

In Section 2, current single-cell downstream analysis methods, pipelines and tools will be discussed. In Section 3, the proposed automatic AI framework will be introduced and explained. In Section 4, the use case will be evaluated with visualisation results. In Section 5, the conclusion will be provided with future work discussions.

II. SINGLE-CELL DOWNSTREAM ANALYSIS METHODS, PIPELINE AND TOOLS

Single-cell RNA sequencing (scRNA-seq) data can be analysed using big data processing and machine learning technologies. These technologies enable the investigation of complex biological questions at the single-cell level that can result in more direct and physiologically relevant and urgently needed observations and understandings. [8] Briefly, the single-cell RNA sequencing process can be broken down into several sequential stages [9], [10]:

- Single-cell isolation can occur from healthy or diseased tissue from any organ of interest. The cell is lysed, which is followed by RNA isolation, purification and quantification, RNA fragmentation, and cDNA generation (reverse transcription uses primers are used to initiate binding to its complementary sequences on the RNA template and serves as a starting point for the synthesis of a new strand ensuring the preservation of original cellular information [11].
- Library-based amplified and tagged cDNA from each cell can be pooled and sequenced [12]. At the end of this stage, large quantities of raw data are prepared.
- Computation algorithms focused on downstream data analysis. The analysis can be used for data reprocessing (quality control), normalisation, feature extraction to clustering, sub-population identification, and understanding gene expression differences across different contexts [13].

In the downstream analysis, quality control refers to the identification of low-quality cells (culture of single cells in droplets, plates, or microfluidic devices can be technically challenging, which can result in the cell undergoing biological stress or even death. On occasion, data from more than one cell can be captured, or even data is recorded where no cell is present at all these undesired variances from the experimental norm are referred to as “low-quality”). This is generally achieved by analyzing the raw data, which is the most critical step for downstream analysis and results in interpretation [14]. Numerous tools (i.e., FASTQC, Kraken, and RNA-SeQC) [15] with different metrics have been developed for quality control, such as the total number of reads detected per cell and the total number of unique genes detected in each sample. These tools provide interfaces allowing researchers to upload the raw data to visualise and process the data with specified metrics and thresholds. For example, “External RNA Controls Consortium Spike-In Controls” can be used to provide information on the sensitivity, specificity, and dynamic range of the datasets by measuring abundances and ratios between spike-in RNAs and endogenous RNAs. This ratio can estimate the total amount of RNA in the captured cells.

The normalisation step is also essential to ensure accurate and reliable downstream analysis. Normalisation approaches not only account for sequencing depth but also account for library sizes. Library sizes vary for many reasons, including natural differences in cell size, variation of RNA capture, and variation in the efficiency of PCR amplification used to

generate enough RNA to create the sequencing library. There are two main approaches to this correction. Many methods use a simple linear scaling to adjust counts such that each cell (row) has about the same total library size. Examples include converting to counts per million (CPM) and closely related methods such as *scrn*. While simple, these approaches do a reasonable job of correcting for differences in library size. Other methods are more complex and helpful in dealing with complex sources of unwanted variation (e.g., for highly heterogeneous populations of cells with different sizes) [18]. The extra function of Removing Unwanted Variation (RUV) is proposed by [19], which adjusts for nuisance technical effects by performing factor analysis on suitable sets of control genes (e.g., ERCC spike-ins) or samples (e.g., replicate libraries).

The feature extraction step focuses on dimensionality reduction which will increase the analysis interoperability (a large set of variables and return a smaller set of components that still contain most of the information in the original dataset) and decrease the analysis complexity. The most popular algorithms are tSNE (t-Distributed Stochastic Neighbour Embedding) [16] and UMAP (Uniform Approximation and Projection) [17]. tSNE combines dimensionality reduction (e.g., PCA) with random walks on the nearest-neighbour network to map high-dimensional data to a 2-dimensional space. UMAP is a non-linear and nondeterministic dimensionality reduction method that requires the random seed to ensure reproducibility. While tSNE optimises for local structure, UMAP tries to balance the preservation of local and global structure [18].

The final step is clustering or classification analysis to understand differences in gene expression. Machine learning algorithms such as K-means, logistic regression, support vector machines, random forests, and neural networks can be applied [20]. Asking bioinformatics scientists and researchers to code solutions step-by-step is not helpful or efficient. Therefore, many pipeline tools are developed such as *Scater* [13] (a pipeline R library to support researchers to have a full programming package on the downstream analysis). Most recently the *devCellPy* [21] (a Python tool that enables automated prediction of cell types across complex annotation hierarchies) and the R code pipeline [22] and *scWizard* (a web application tool for specifying the template of the downstream analysis) [23] have been utilised.

However, these pipelines or the full stack of development packages still require high-level knowledge of coding with a specific programming language. Therefore, reusing, exchanging, and sharing the experimental environments and research outcomes in cross-disciplinary collaboration are very challenging [24]. In addition, comparing many different algorithms to find the best one for processing data and analysis is extremely time-consuming and requires very specialised knowledge [4], [25]. For instance, independent research groups have not extensively used Deep Learning (DL) algorithms in their biological studies due to a lack of expertise and robust computation resources [26]. Finally, cutting-edge technologies will be delayed in application because of the high bar of programming. Thus, an automated framework will be the key

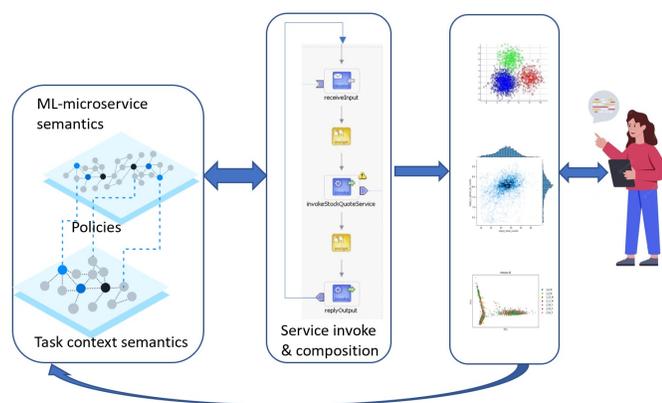


Fig. 1. Platform Architecture

to addressing these challenges [27]. The paper proposes a framework based on advanced knowledge graphs, microservices, and knowledge-based AI technologies. This framework aims to separate algorithm development and analysis tasks for different disciplinary researchers. In the end, the platform will be able to automatically create a downstream analysis pipeline for bio-scientists.

III. THE PROPOSED FRAMEWORK ARCHITECTURE AND PROTOTYPE DEVELOPMENT

The proposed framework consists of three layers as shown in Figure 1:

- Semantic layer that provides metadata descriptions of the analysis task context, policies, and microservices registered in the platform. The task context simply specifies the inputs, analysis task, domain, and desired output data. Each microservice only has one function to do a specific task that can be performed in a different stage of the analysis. The stages are normally data translation and loading, data normalisation, data processing, and data analysis.
- Automation layer that selects and coordinates microservices to create a pipeline dynamically for completing the task. The task can be a simple task handled by one single microservice. However, most of the functions need to compose several services together dynamically. The automation performs selections and coordination through semantic reasoning and context-based reinforcement learning.
- The output and interaction layer that releases the results of the analysis process which can be single microservice outputs or a pipeline's outcomes produced at each stage of the analysis. The researchers can interact with the system at any stage to suggest ratings and provide further context. The interaction data will be fed back to the semantic layers to enhance the policies.

Service Registration		
Name/Space	Microservice Name	Description
<input type="text" value="the service namespace e.g http://serviceservice.io"/>	<input type="text" value="the service name"/>	<input type="text"/>
Microservice Type	Dependencies	Framework
<input type="text" value="data loading, data streaming, data transfer, data s"/>	<input type="text" value="libraries needed e.g [stream..., pandas..., numpy]"/>	<input type="text" value="programming frameworks needed, separate with a s"/>
Turns on the Usage of GPU or not	Pipeline	Model Format
<input type="text" value="True"/>	<input type="text" value="the microservices composition process of this service"/>	<input type="text" value="the model save formation e.g. h5 that can be invok"/>
Service Category	Input/Output	Output/Op
<input type="text" value="Linear Regression"/>	<input type="text" value="Datafile.csv"/>	<input type="text" value="Datafile.csv"/>
Contributor	License	
<input type="text" value="organisation or individual or groups who developed"/>	<input type="text" value="License if Available"/>	

Fig. 2. Microservice registration interface

A. Semantic Knowledge Graph schema

There are four types of schema defined using OWL (Web Ontology Language) [28]. OWL is a knowledge graph ontology design standard.

- The microservice ontology contains namespace, identity, input data, output data, domain, dependency, purpose, and ML category, as well as other properties to able, create the dynamic invocation settings (see Figure 2).
- The task context ontology includes task identity, task input data, task desired output data, and domain.
- The solution pipeline ontology defines a workflow created by having one or composing multiple microservices to achieve the desired output.
- The policies ontology defines scores for each service against each task context with a default score of 0. This means that the platform will learn the procedures and try to remember the best solution by evaluating all possible microservices or combinations.

B. Dynamic microservices selection and composition

We implemented five engines to deal with microservices selection, invocation, optimisation, composition, and policy learning. The process of dynamically creating a solution is represented in Figure 3.

The analysis task context is the input to trigger the automatic process and the search engine then starts to create a SPARQL (a knowledge graph query language) [29] query to match semantic compatible microservices that can produce the desired output. The outcomes from the search engine can be either, a single result (a compatible single microservice or pipeline found), multiple results, or no result. The first situation is simply to complete the task and feed the results to the task requester, then the requester can provide the score as feedback to the policy engine. For the second situation, a queue containing all possible compatible results is created to allow the invocation engine to invoke them one by one to feed the output to the optimisation engine that can select the best solution through quality and performance evaluation as well as run-time feedback from the task requester. For the no-matching condition, the composition engine starts to try to compose a sequence of microservice services to complete the task by relaxing some context-searching criteria. The composition is

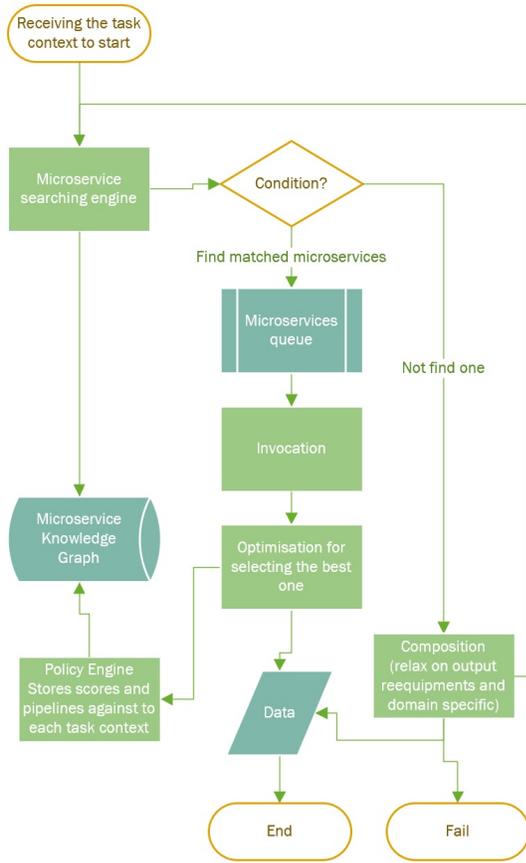


Fig. 3. Automatic data processing and analysis engines

completed when the relaxed criteria are no longer required after one or some other microservices produced a mediation outcome to fill the gap. The composition process may also succeed or fail. The results will be recorded through the policy engine to remember the successful microservice, pipeline of them, or no solution outcomes against a task context. In the future, a similar task will be performed faster by reusing the whole solution or part of the solution.

C. Interfaces for interaction

The web interfaces are developed in the framework to support interactions. AI microservices can be registered and shared by researchers or AI engineers (see Figure 2). Researchers can then use the task interface to specify the analysis task for asking the framework to provide the best solution based on the knowledge about the registered microservices (see Figure 4). Whenever a step is completed in the process towards the goal, the output can be presented to the researcher for immediate feedback to support the next steps or overall input to the solution (see Figure 7). All the feedback will contribute to the task context policies that improve the output of the framework.

IV. USE CASE DEMONSTRATION

In this section, we use a clustering analysis case study to highlight how the proposed framework can solve a real-world

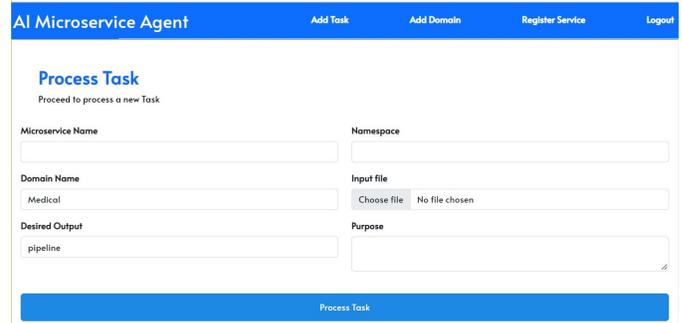


Fig. 4. Task specification interface

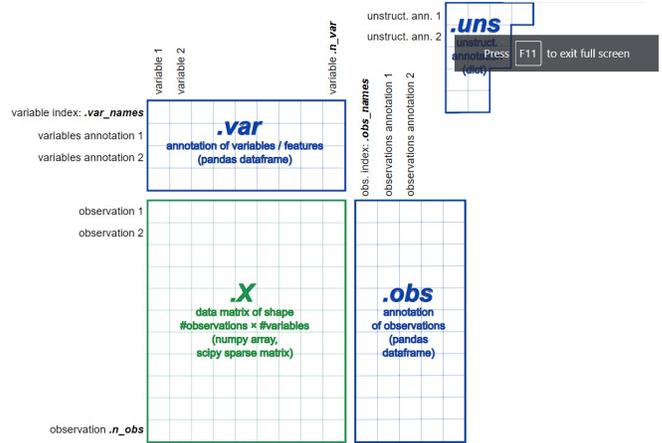


Fig. 5. AnnData Structure

downstream single-cell data analysis task. The clustering analysis task works on a mouse brain single-cell RNASeq dataset. The dataset is publicly available through a workshop tutorial at [30]. There are five sequential processing and analysis steps:

- 1) **Data semantic transforming and loading:** For instance, applying AnnData structure [31], where AnnData stores observations (samples) of variables/features in the

```
@prefix ns1: <http://aimicroservice.derby.ac.uk/> .
ns1:genQualityControl a ns1:Bioinformatic_genQualityControl ;
ns1:category ns1:Bioinformatics ;
ns1:contributor <https://www.derby.ac.uk/staff/hongqing-yu/> ;
ns1:dependency "matplotlib"@en,
  "pandas"@en,
  "scanpy"@en,
  "seaborn"@en ;
ns1:description "https://scanpy.readthedocs.io/en/stable/"@en ;
ns1:formate "py"@en ;
ns1:framework "annData_qualityControl"@en ;
ns1:input [ ns1:parameter [ ns1:iocategory ns1:brain_raw ;
  ns1:iodatatype ns1:h5ad ;
  ns1:pid "0"@en ] ] ;
ns1:licence <https://en.wikipedia.org/wiki/Free-software_license> ;
ns1:output [ ns1:parameter [ ns1:iocategory ns1:brain_qc ;
  ns1:iodatatype ns1:h5ad ;
  ns1:pid "0"@en ] ] ;
ns1:uri ns1:genQualityControl .
```

Fig. 6. Quality control microservice semantic description

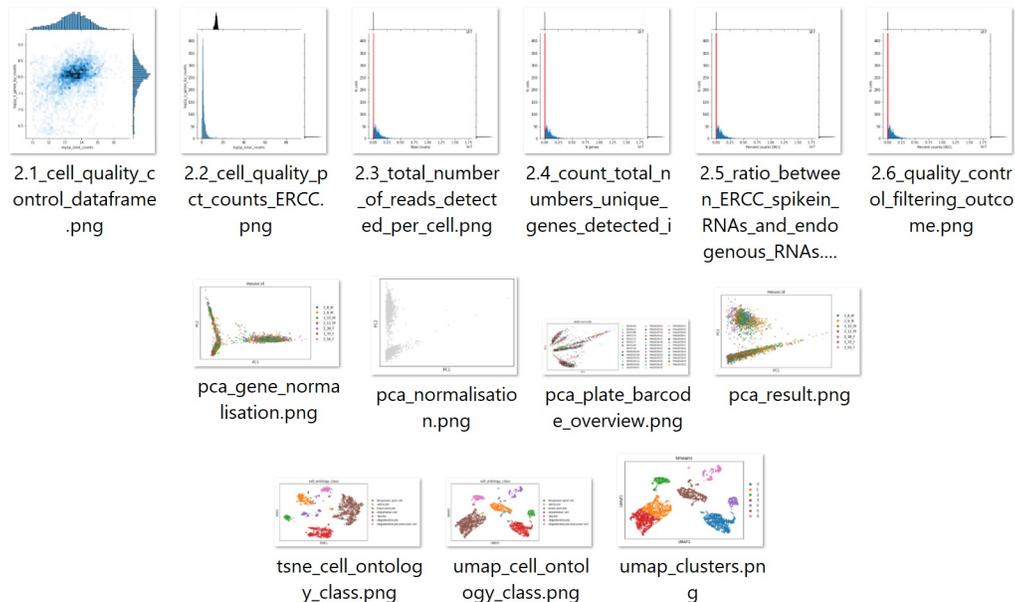


Fig. 7. Visualisations of analysis steps

rows of a matrix (see Figure 5).

- 2) **Data quality control:** This aims to find and remove the poor quality cell observation data which were not detected in the previous processing of the raw data. The low-quality cell data may potentially introduce analysis noise and obscure the biological signals of interest in the downstream analysis.
- 3) **Data normalisation:** Dimensionality reduction and scaling of the data. Biologically, dimensional reduction is valuable and appropriate since cells respond to their environment by turning on regulatory programs that result in the expression of modules of genes. As a result, gene expression displays structured co-expression, and dimensionality reduction by the algorithm such as principle component analysis can group those co-varying genes into principle components, ordered by how much variation they explained.
- 4) **Data feature embedding:** Further dimensionality reduction using advanced algorithms, such as t-SNE and UMAP. They are powerful tools for visualising and understanding big and high-dimensional datasets.
- 5) **Clustering analysis:** Group cells into different clusters based on the embedded features.

Based on the above five steps, we developed seven microservices which include AnnData loading, data quality control, normalisation services (PCA+CPM algorithm), two feature embedding services (t-SNE and UMAP), and clustering services (K-mean clustering and Louvain graphical clustering algorithms).

The microservices were semantically registered into the framework through the interface. Figure 6 depicts an example of quality control microservice semantic description in the knowledge graph repository.

With all the microservices registered, researchers can start expressing the analysis task to stop, interact and provide feedback at any stage during the process of automatically creating the solution. The researchers can also see visualisations of outputs produced by different steps (see Figure 7). Therefore, researchers can provide preferences for selecting microservices if there are options.

A realistic example is that a researcher can specify a clustering task applied to the mouse brain single-cell RNASeq dataset. The framework will first try to see if a single microservice can complete this task. The answer is 'no' because no semantic-matched microservice can take the RNASeq CSV input and provide the clustering output. At this juncture, the microservice that can take the RNASeq CSV will be invoked to process the data into the next step with the output of AnnData. If there are multiple choices in the composition sequence, all possibilities will be invoked to run unless the previous knowledge in the policies has a priority. The possibilities have multiple solutions at the end for researchers to analyse for giving professional feedback to the system. The feedback will help greatly with the knowledge graph policies. For example, suppose the researcher gives feedback to the system that UMAP is the better embedding method than t-SNE but has no priority on the clustering methods. In that case, the framework will produce two possible clustering results shown in Figure 8.

V. CONCLUSION AND FUTURE WORK

The potential of single-cell research, along with ML and AI technologies, to address critical biomedical and disease classification and clustering issues and facilitate comprehension in the near future is substantial [32], [33]. Our research identified the current limitations of reusing, exchanging, sharing,

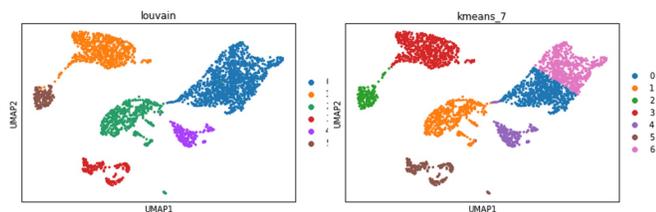


Fig. 8. Two clustering outcomes from automatic process

applying the most advanced technologies and automating the experimental environments and outcomes in cross-disciplinary research collaboration. Therefore, we proposed an AI automation framework that can semantically share implemented ML algorithms or AI models for a general purpose. Possible solutions can be automatically generated through interactions with researchers.

Our future work will increase the general informatics purposed ML algorithms and AI model developments and registration. More downstream analysis tasks can be tested and evaluated.

REFERENCES

- [1] Perkel, J., (2021). Single-cell analysis enters the multiomics age. *Nature*. 595. 614-616. <https://doi.org/10.1038/d41586-021-01994-w>.
- [2] Stuart, T. and Satija, R., Integrative single-cell analysis. *Nat Rev Genet* 20, 257–272 (2019). <https://doi.org/10.1038/s41576-019-0093-7>
- [3] Pasquini, G. Eduardo Rojo Arias, J., Schäfer, P. and Busskamp, V., Automated methods for cell type annotation on scRNA-seq data, *Computational and Structural Biotechnology Journal*, Volume 19, 2021, Pages 961-969, ISSN 2001-0370, <https://doi.org/10.1016/j.csbj.2021.01.015>.
- [4] Yuan, G. C., Cai, L., Elowitz, M. et al., Challenges and emerging directions in single-cell analysis. *Genome Biol* 18, 84 (2017). <https://doi.org/10.1186/s13059-017-1218-y>
- [5] Chen, G., Ning, B. and Shi, T., Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front Genet*. 2019 Apr 5;10:317. doi: 10.3389/fgene.2019.00317. PMID: 31024627; PMCID: PMC6460256.
- [6] Raimundo, F., Meng-Papaxanthos, L., Vallot, C. and Vert, J., Machine learning for single-cell genomics data analysis, *Current Opinion in Systems Biology*, Volume 26, 2021, Pages 64-71, ISSN 2452-3100, <https://doi.org/10.1016/j.coisb.2021.04.006>.
- [7] Ma, Q., Xu, D., Deep learning shapes single-cell data analysis. *Nat Rev Mol Cell Biol* 23, 303–304 (2022). <https://doi.org/10.1038/s41580-022-00466-x>
- [8] Zhang, Z. and et al., Critical downstream analysis steps for single-cell RNA sequencing data, *Briefings in bioinformatics* (2021): n. pag.
- [9] Haque, A., Engel, J., Teichmann and S. A. et al. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med* 9, 75 (2017). <https://doi.org/10.1186/s13073-017-0467-4>
- [10] Slovin, S., Carissimo, A., Panariello, F., Grimaldi, A., Bouché, V., Gambardella, G. and Cacchiarelli, D., Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. *Methods Mol Biol*. 2021, 2284:343-365.
- [11] Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. and Taipale, J., Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*. 2011 Nov 20;9(11):72-4. doi: 10.1038/nmeth.1778. PMID: 22101854.
- [12] van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thernes, C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;30:418–26.
- [13] McCarthy, D. J., Campbell, K. R., Lun, A. T. L. and Wills, Q. F., Scater: pre-processing, quality control, normalisation and visualization of single-cell RNA-seq data in R, *Bioinformatics*, Volume 33, Issue 8, 15 April 2017, Pages 1179–1186, <https://doi.org/10.1093/bioinformatics/btw777>
- [14] Li, X., Nair, A., Wang, S. and Wang, L., (2015). Quality Control of RNA-Seq Experiments. In: Picardi, E. (eds) *RNA Bioinformatics. Methods in Molecular Biology*, vol 1269. Humana Press, New York, NY. <https://doi.org/10.1007/978-1-4939-2291-8-8>
- [15] Bacher, R. and Kendziorski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol* 17, 63 (2016). <https://doi.org/10.1186/s13059-016-0927-y>
- [16] N. Rogovschi, J. Kitazono, N. Grozavu, T. Omori and S. Ozawa, t-Distributed stochastic neighbor embedding spectral clustering, 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 1628-1632, doi: 10.1109/IJCNN.2017.7966046.
- [17] McInnes, L. and Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv*, abs/1802.03426.
- [18] Sharma, A., Single-cell RNASeq data from Mouse Brain, 2021, <https://www.kaggle.com/datasets/aayush9753/singlecell-rnaseq-data-from-mouse-brain/discussionSingle-cell>
- [19] Risso, D., Ngai, J., Speed, T. and et al., Normalisation of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 32, 896–902 (2014).
- [20] Le, H., Peng, B., Uy, J., Carrillo, D., Zhang, Y. and Aebermann, B. D., Scheuermann RH. Machine learning for cell type classification from single nucleus RNA sequencing data. *PLoS One*. 2022 Sep 23;17(9):e0275070. doi: 10.1371/journal.pone.0275070. PMID: 36149937; PMCID: PMC9506651.
- [21] Galdos, F. X., Xu, S., Goodyer, W. R. and et al., devCellPy is a machine learning-enabled pipeline for automated annotation of complex multilayered single-cell transcriptomic data. *Nat Commun* 13, 5271 (2022). <https://doi.org/10.1038/s41467-022-33045-x>
- [22] Chen, Y., Pal, B., Lindeman, G.J. and et al. R code and downstream analysis objects for the scRNA-seq atlas of normal and tumorigenic human breast tissue. *Sci Data* 9, 96 (2022). <https://doi.org/10.1038/s41597-022-01236-2>
- [23] Wei, J., Xie, Q., Qu, Y., Huang, G., Chen, Z. and Du, H., scWizard: A web-based automated tool for classifying and annotating Single-cells and downstream analysis of single-cell RNA-seq data in cancers, *Computational and Structural Biotechnology Journal*, Volume 20, 2022, Pages 4902-4909, ISSN 2001-0370, <https://doi.org/10.1016/j.csbj.2022.08.028>.
- [24] Hodzic, E., Single-cell analysis: Advances and future perspectives. *Bosn J Basic Med Sci*. 2016 Nov 10;16(4):313-314. doi: 10.17305/bjbm.2016.1371. PMID: 27320288; PMCID: PMC5136769.
- [25] Menon, S., Lui, V.C.H. and Tam, P.K.H., 2021. Bioinformatics tools and methods to analyze Single-cell RNA sequencing data. *International Journal of Innovative Science and Research Technology*,(IJISRT), 6(8), pp.282-288.
- [26] Ma, Q., Xu, D. Deep learning shapes single-cell data analysis. *Nat Rev Mol Cell Biol* 23, 303–304 (2022). <https://doi.org/10.1038/s41580-022-00466-x>
- [27] Saliba A. E., Westermann A. J., Gorski S. A., Vogel J., Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res*. 2014 Aug;42(14):8845-60. doi: 10.1093/nar/gku555. Epub 2014 Jul 22. PMID: 25053837; PMCID: PMC4132710.
- [28] Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneijder, P. and Stein, L. A., (2004). OWL Web Ontology Language Reference (Recommendation). World Wide Web Consortium (W3C).
- [29] Prud'hommeaux, E. and Seaborne, A., SPARQL Query Language for RDF , W3C Recommendation, 2008.
- [30] Luecken, Malte D. and Theis, Fabian J., Current best practices in single-cell RNA-seq analysis: a tutorial, *Journal of Molecular Systems Biology*, Volume 15, 2019, <https://doi.org/10.15252/msb.20188746>. Dataset download url: <https://www.singlecellcourse.org/>
- [31] Cannoodt, R., (2022). `anndata: 'anndata'` for R. <https://anndata.dynverse.org>, <https://github.com/dynverse/anndata>.
- [32] Chen, Y., Pal, B., Lindeman, G.J. and et al., R code and downstream analysis objects for the scRNA-seq atlas of normal and tumorigenic human breast tissue. *Sci Data* 9, 96 (2022). <https://doi.org/10.1038/s41597-022-01236-2>
- [33] Dohmen, J., Baranovskii, A., Ronen, J. and et al., Identifying tumor cells at the single-cell level using machine learning. *Genome Biol* 23, 123 (2022). <https://doi.org/10.1186/s13059-022-02683-1>

Machine Learning in the Identification of Key Residues of Variants and Polymorphisms in the Interaction of ACE2 Proteins with Spike of SARS-CoV2

Ana Carolina Damasceno Sanches, Ana Luisa Rodrigues de Avila, Levy Bueno Alves and Silvana Giuliatti

Department of Genetics, Ribeirao Preto Medical School

University of Sao Paulo, USP

Ribeirao Preto, Brazil

e-mail: silvana@fmrp.usp.br

Abstract—The binding affinity between the Spike protein and the angiotensin-converting enzyme 2 receptor (ACE2) is one of the main determining factors in the replication rate of Severe Acute Respiratory Syndrome of Coronavirus-2 that directly affects the clinical condition of the patient. The presence of multiple variants indicates a high mutation rate of the virus. Furthermore, genetic variations within the coding regions of ACE2 can impact the susceptibility, severity, and progression of the disease. However, the effect of these mutations on the stability and affinity of the Spike-ACE2 interaction is not well understood. To gain insight into this interaction, molecular dynamics simulations are used. Although these simulations produce a large amount of data, they do not make easy to identify residues that play a significant role in the interaction between the proteins. To overcome this issue, we combined molecular dynamics simulations and supervised machine learning techniques to identify the residues that have the most impact on the interaction and dynamics of the complexes. The molecular dynamics simulations showed slight variations in complex trajectories, but highlighted key residues and loop region residues. Despite stable behavior among variants with only minor differences, the machine learning methods identified critical residues in ACE2 and Spike proteins that can affect virus-host interaction.

Keywords—COVID-19; Bioinformatics; Molecular Docking; Polymorphism; Variants.

I. INTRODUCTION

On March 11, 2020, the World Health Organization characterized COVID-19 [1] as a pandemic, an infectious disease caused by the Severe Acute Respiratory Syndrome of Coronavirus-2 (SARS-CoV-2). To date, November 2022, more than 630 million cases have been confirmed, including more than 6.6 million deaths globally. In Brazil alone, there are more than 35 million cases with almost 690,000 deaths [2]. COVID-19 is a respiratory disease, transmitted by the epithelial cells of the lung through aerosols, which can lead from mild viral pneumonia to Acute Respiratory Distress Syndrome, and in even more serious cases leading to multiple organ failure [3]. It mainly affects individuals with comorbidities and/or some type of immunosuppression. Some people develop the severe form of COVID-19, while others are asymptomatic [3][4]. The entry of the virus into the cell is one of the most important processes in viral

infection, being the target in the development of vaccines and drugs. The invasion of SARS-CoV-2 into host cells depends on the interaction of the Spike structural protein with the human protein, present in the cell membrane, angiotensin-II converting enzyme [5] and variants of this protein have been associated with susceptibility to SARS-CoV-2 [3][4].

SARS-CoV-2 has a high probability of mutating and adapting better to the environment [4]. The current Variant of Concern (VOC) is Omicron (B.1.1.529 - several countries), prior to this, also classified as VOC: Alpha (B.1.1.7 - United States), Beta (B.1.351 - Africa do Sul), Gamma (P.1 - Brazil) and Delta (B.1.617.2 - India) [1]. P2 variant (or Zeta variant) (B.1.1.28.2) was detected in the city of Rio de Janeiro in October 2020. The mutations suffered by SARS-CoV-2 observed in its variants, as well as the polymorphisms observed in the ACE2 protein, raise questions such as whether genetic variability of the virus and the host could explain the different degrees of severity in cases of infection. How these mutations contribute to improving the stability and affinity between Spike-ACE2 complexes is not a process fully understood.

Molecular Dynamics simulations have been used to assess the stability and affinity between complexed structures. The trajectories resulting from these simulations generate large amounts of data from thousands of atoms at each time interval. The stability of the complex is analyzed by calculating the root-mean-square deviation and also by the number and type of contact between the structures. However, the high-dimensional nature and noisy output of the simulations make it extremely difficult to extract meaningful features from the trajectories, thereby hindering a deeper understanding of molecular processes

Machine learning techniques are employed to analyze vast data sets. These methods assist in identifying key differences between the trajectories obtained from molecular dynamics simulations. Fleetwood et al. [6] demonstrated the usefulness and potential of machine learning techniques in comprehending biomolecular processes by applying both supervised and unsupervised techniques to three different biological systems. Inspired by this work, we utilized molecular dynamics simulations to evaluate the stability of complexes and applied supervised machine learning techniques using the resulting trajectories as input data to investigate the effect of genetic variability in SARS-CoV-2 and ACE2 polymorphisms on the interaction region between these proteins.

II. MATERIAL AND METHODS

In this section, we will outline the methods employed to perform molecular dynamics simulations and implement machine learning architectures.

A. Molecular Dynamics

The tertiary structure of the complex Spike and ACE2 (PDB ID: 6LZG) was obtained from the Protein Data Bank [7]. Modeller software v9.23 [8] was used to fill the missing atoms.

GROMACS package version 2019.3 [9] was used in the MD simulations of complexes. The force field used was CHARMM36 [10]. The molecules were solvated with TIP3P water molecules and neutralized by adding the appropriate number of Na+Cl ions considering the ionic concentration of 0.15 M. The energy minimisation was performed using the steepest descent method with a maximum force of 1000 KJ/mol.nm. After minimization, the systems were equilibrated in two stages: a canonical NVT set followed by an isothermal-isobaric NPT set. The NVT equilibrium was performed with a constant temperature of 300 K for 500 ps. The NPT equilibrium was performed with a constant pressure of 1 bar and a constant temperature of 300 K for 500 ps. The production step was conducted at 300 K for 100 ns and the trajectories were saved every 10 ps. Four complexes ACE2-Spike complexes were analysed: ACE2-Spike (wild) and 3 ACE-Spike (Omicron, Delta and P2 – Zeta variant).

B. Machine Learning

Based on Fleetwood et. al [6], we employed molecular dynamics trajectories as input for supervised ML techniques. To reduce the influence of a single model and enhance the stability of our results, we utilized two differing supervised machine learning classification strategies: Multilayer Perceptron (MLP) and Random Forest (RF). These methods were used to identify residues that most significantly contribute to the difference in the dynamic behavior between the complexes (Fig. 1). A multilayer perceptron is a type of artificial neural network has multiple layers between input and output layers. Meanwhile, Random Forest is an ensemble learning technique that is used for classification by building many decision trees and finding the mode of the classes of each tree. We chose to use both RF and MLP because they are powerful and commonly-used supervised machine learning algorithms. RF excels at performing both regression and classification tasks and is well-known for its robust performance and handling of noisy and missing data. MLP, a feedforward neural network, can handle regression and classification problems, and is frequently used for complex, non-linear relationships

The input features for these algorithms include the contact distances between ACE2 residues and Spike. These distances were calculated as the minimum distance between the heavy atoms of residues in the interaction region. Only distances less than 15 Å were considered in forming our

dataset. The values were then inverted, normalized and used as inputs.

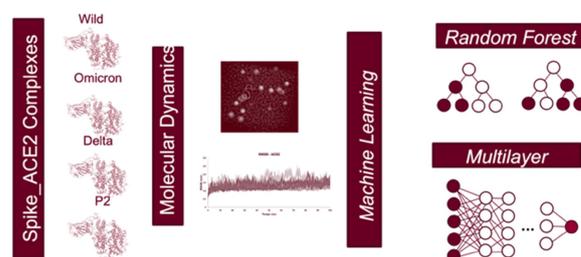


Figure 1. Flowchart of the Machine Learning methods used on this study.

The MLP was implemented using the open-source machine learning library Scikit-learn in Python [11]. We also used the data analysis and manipulation library Pandas [12], and the numerical computing library NumPy [13]. Scikit-learn is a widely-used, well-documented, and efficient machine learning library that provides quick prototyping and testing.

We employed 8 hidden layers with 100, 75, 50, 40, 30, 20, 10, and 5 neurons respectively, with ReLU activation. ReLU is a popular activation function in deep learning that is known for its effectiveness. The labels were one-hot encoded to represent categorical data numerically. The training process used the Adam optimizer [14] to adjust the node weights. This optimizer is frequently used due to its demonstrated efficacy.

We created the first profile by building a correlation matrix for training and testing. Four additional profiles were generated through bootstrapping and features with strong correlation were discarded using a 0.9 threshold. As a result, 5 profiles were obtained with 1828, 1907, 1925, 1909 and 1934 features respectively, each with 40 thousand frames. The MLP was trained with each of these profiles, resulting in 10 total MLPs. We used a train-test split to evaluate the performance of the ML algorithms, with 80% of the data in the training set and 20% in the test set.

Important features for classification were determined using Layer-Wise Relevance Propagation (LRP) [15] with the LRP-0 rule. LRP assigns relevance scores to input features, making it possible to visualize which inputs have the most impact on a specific prediction made by the model. This enhances transparency and confidence in the decision-making of neural networks

Our Random Forest model utilized the Gini impurity coefficient, which ranges from 0 to 1, with 0 indicating a pure split and 1 representing maximum impurity. The aim was to choose splits that would lower Gini impurity, resulting in more homogeneous class distribution in the tree's leaves. RF The RF classifier uses an internal bootstrapping process to produce consistent profiles. The model consisted of 100 decision trees, with 3201 features

and 40 thousand frames. The one-versus-the-rest method was employed to calculate feature importance, a strategy in multi-class classification that plits the problem into several binary classification problems. RF was implemented using the Scikit-learn library.

III. RESULTS AND DISCUSSION

The outcomes achieved at each step of our work will be detailed in the subsequent sub-sections.

A. Molecular Dynamics

We sought differences in the interactions between SARS-CoV-2 variants and the ACE2 protein through 100 ns molecular dynamics simulations for each complex. The simulation data was used to compute Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF). Fig. 2A and fig. 2B show RMSD values for ACE2 and Spike proteins, respectively.

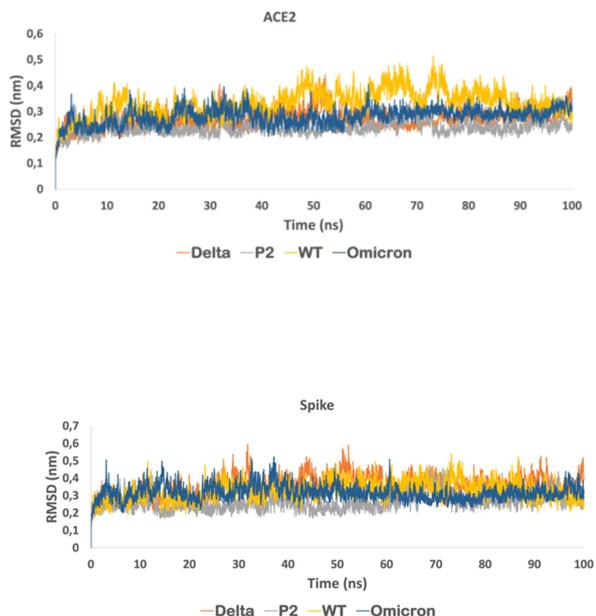


Figure 2. Analysis of the trajectories obtained in the MD simulation. (A) RMSD of ACE2 (B) RMSD of Spike.

Results demonstrate stability in the ACE2 protein for all complexes at around 10ns, with similar RMSD values ranging from 0.2 to 0.4 nm.

The RMSF analysis of the ACE2 trajectory (Fig. 3A) showed no significant fluctuations, limited, wich were limited to loop regions.

The Spike protein (Fig. 3B) showed that the Lys444 residue of the delta variant had the highest fluctuation peak of 0.20 nm, followed by the omicron variant (0.16 nm), P2 variant (0.16 nm), and wild-type (0.14 nm). Lys444 is located close to Gly446, Tyr449, and Gln498, which have polar interactions with ACE2, according to a study by Sironi et al. [16].

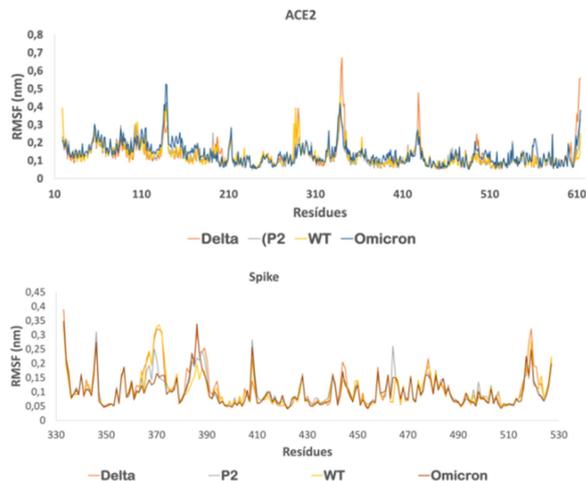


Figure 3. Analysis of the trajectories obtained in the MD simulation. (A) RMSF of ACE2 (B) RMSF of Spike.

The Tyr449 residue is situated near to Leu452, which was mutated to arginine in the delta variant. Other residues with high fluctuation peaks are located in loop regions.

B. MLP

Table 1 shows the five most significant pairs for each complex. The residue importance values for each pair were determined by finding the average LRP-0 value assigned to these pairs in the generated MLPs. Key residues responsible for differences in binding between Spike variants and ACE2 have been identified. Some of these were previously noted in previous studies.

TABLE I. RESIDUES IMPORTANCE OBTAINED FROM MLP

Variant	MLP	
	Residue Pair (ACE, Spike)	Importance Value
Wild	(SER106, GLY485)	1.0
	(VAL107, PHE486)	0.99
	(GLN89, SER477)	0.98
	(SER19, PRO479)	0.89
	(ALA71, GLU484)	0.82
Delta	(ASP30, GLU484)	1.0
	(GLN24, LYS417)	0.71
	(GLY352, ARG408)	0.68
	(ALA65, SER443)	0.62
	(ASN33, GLN498)	0.60
Omicron	(GLU329, SER438)	1.0
	(GLN42, SER349)	0.96
	(TYR381, GLY502)	0.92
	(GLY352, ASN448)	0.91
	(GLY354, GLY504)	0.9
P2	(PRO321, ARG403)	1.0
	(SER19, ASN477)	0.92
	(SER19, PRO479)	0.87
	(GLN325, SER371)	0.84
	(GLU37, THR415)	0.81

The analysis of the results highlights the role of key ACE2 residues GLN24, GLN42, GLN325, GLU329, and

GLY354 in interaction with protein S. Moreover, ACE2 residue SER19, which was commonly seen among the pairs, is also important. A mutation in this residue S19 to P increased the interaction between ACE2 and Spike protein. However, mutations in residues ASN33 (N33I) and GLY352 (G352V) were found to reduce this interaction [17].

Our results highlight the key residues in the Spike protein that can contribute to variations in binding between Spike variants and ACE2. The mutation of the LYS417 to K417N, increases virus transmissibility. The SER477 residue (S477N mutation) enhances binding affinity. The GLU484 residue, when mutated to E484K, has been linked to antibody resistance [4]. The only exception is the simultaneous presence of the (SER19, PRO479) pair in both the Wild and P2 variants, as no other pair of residues showed were significance across other variants.

C. RF

Table 2 displays the top five residues that were determined by the Random Forest model, based on their Gini importance values.

TABLE II. RESIDUES IMPORTANCE OBTAINED FROM RF.

Variant	RF	
	Residue Pair (ACE, Spike)	Importance Value
Wild	(SER19, VAL483)	1.0
	(SER19, CYS488)	0.95
	(SER19, CYS480)	0.60
	(SER44, TYR505)	0.52
	(SER19, GLN474)	0.52
Delta	(ALA36, ASN501)	1.0
	(GLY66, ASN501)	0.98
	(ALA342, THR500)	0.93
	(ASN103, TYR505)	0.71
	(LYS68, ASN501)	0.64
Omicron	(ALA25, ASN417)	1.0
	(GLN24, ASN417)	0.98
	(ILE21, ASN417)	0.97
	(LYS353, ARG498)	0.94
	(THR27, ASN417)	0.88
P2	(SER106, LYS484)	1.0
	(SER19, CYS480)	0.90
	(SER105, ASN487)	0.87
	(GLY104, ASN487)	0.80
	(SER105, LYS484)	0.79

The residues pairs identified by the Random Forest model differed from those identified by the MLP model. However, some residues were identified by both methods. Several of these residues have been previously reported, including TYR505 in the Spike protein, whose mutation can increase transmission [4], and ARG498 in the Omicron variant, which leads to increased binding affinity with ACE2 [18]. SER19, LYS353, and THR27 are crucial residues in ACE2 [17]. SER19 was found repeatedly among pairs and variants. The exception was the (SER19, CYS480) pair, which was present in both the Wild and P2 variants, but no other residue pair was present in multiple variants.

IV. CONCLUSIONS AND FUTURE WORK

The interaction between the Spike and ACE2 proteins is crucial in determining the replication rate of SARS-CoV-2 and affects the progression of the disease in infected patients. SARS-CoV-2 exhibits a high mutation rate, as evidenced by the emergence of various variants over the past two years. Polymorphisms in the coding regions of ACE2 may impact a patient's susceptibility to the disease, as well as its severity, and clinical outcome. However, the impact of mutations and polymorphisms on the stability and interaction between the SARS-CoV2-ACE2 complex is not yet fully understood.

In our work, we combined molecular dynamics simulations and machine learning techniques to examine the interaction between SARS-CoV-2 variants and human ACE2. The simulations provided insight into the protein complex interaction, while ML methods identified important residues in the binding region.

Our molecular dynamics simulations showed stability similarities among the variants. The ACE2 protein complex with Spike-Wild showed slightly lower stability, as indicated by RMSD values, compared to the SARS-CoV-2 variant complexes. This aligns with the expectation that mutations in the Spike interaction region increase stability. The ACE2 protein in the wild-type complex is therefore more flexible and less stable. The Spike protein in the Delta variant had slightly higher RMSF values, with a peak at Tyr444 near key residues that interact with ACE2, including Tyr449 near the L452R mutation. Replacing the hydrophobic Leucine with the polar Arginine may enhance intra- and intermolecular interactions.

We achieved an accuracy score of 1 and loss values less than 0.005 for both machine learning methods using the test set. High accuracy and low loss on test data suggest that the model is performing well, not guarantee that the model is not overfitting. Further evaluation using other data sources, such as cross-validation, is needed to determine if overfitting is present.

The ML approaches successfully identified key residues from both proteins responsible for differences in binding region, some of which have been previously reported in the literature. This demonstrates that our method was able to identify residues that significantly contribute to the distinction between virus and host interaction due to mutations in Spike (variants) and ACE2 polymorphisms.

Our study shows that machine learning can simplify the complexity of virus-host interactions by reducing dimensionality and identifying crucial residues. Our findings indicate that there may be additional important residues beyond those previously considered, which can impact the interaction between Spike and ACE2 proteins. These residues could account for differences in stability and affinity, leading to varying levels of susceptibility to SARS-CoV-2 and resulting in varying degrees of disease severity. In our work, we aim to gain a deeper understanding of the relationship between mutations and the affinity between

Spike-ACE2 by not only exploring other variants, but also incorporating various machine learning methods.

ACKNOWLEDGMENT

Fundacao de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for Sponsorship (2022/02782-5; 2022/03038-8) and Superintendence of Technology and Information at USP for the High-Performance Computing (HPC) Computing Resources.

REFERENCES

- [1] WHO Director General's Speeches. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. WHO Director General's speeches, n. March, p. 4, 2020. Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> 2023.01.31
- [2] WHO. Tracking SARS-CoV-2 variants. Who. Disponível em: Available from: <https://www.who.int/activities/tracking-SARS-CoV-2-variants>. 2023.01.31
- [3] S. Choudhary, K. Sreenivasulu, P. Mitra, S. Misra and P. Sharma. "Role of genetic variants and gene expression in the susceptibility and severity of COVID-19". *Annals of Laboratory Medicine*, vol. 41, no. 2, pp. 129–138, 2020.
- [4] J. Zepeda-Cervantes et al. "Implications of the Immune Polymorphisms of the Host and the Genetic Variability of SARS-CoV-2 in the Development of COVID-19." *Viruses*, vol. 14, no 1, pp. 1-34, 2022.
- [5] R. Peng, L-A. Wu, Q. Wang, J. Qi and G. F. Gao " Cell entry by SARS-CoV-2." *Trends in Biochemical Sciences*, vol. 46, no.10, pp 848–860, 2021
- [6] O. Fleetwood, M. A. Kasimova, A. M. Westerlund and L. Delemotte. "Molecular Insights from Conformational Ensembles via Machine Learning." *Biophysical Journal*, vol. 118, no. 3, pp. 765–780, 2020.
- [7] H. M. Berman et al. "The Protein Data Bank", *Nucleic Acids Research*, vol. 28, no 1, pp. 235-242, 2000.
- [8] A. Sali and T. L. Blundell "Comparative protein modelling by satisfaction of spatial restraints", *J. Mol. Biol.*, vol. 234, no. 1, pp. 779-815, 1993.
- [9] M. Mohammadi, M. Shayestehpour and H. Mirzaei "The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines." *Brazilian Journal of Infectious Diseases*, vol. 25, no. 4, pp. 101606, 2021.
- [10] J. Huang and A. D. MacKerell "CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data." *J. Comput. Chem.*, vol. 34, no 1, pp. 2135-2145 2013.
- [11] Scikit-learn: Machine learning in Python. Available from <https://scikit-learn.org/stable/index.html>. 2023.01.31
- [12] Pandas-Python Data Analysis Library. Available form <https://pandas.pydata.org/>. 2023.01.31
- [13] Numpy. Available from <https://numpy.org/>. 2023.01.31
- [14] D. P. Kingma and J. Ba. 2014. Adam: a method for stochastic optimization. arXiv, arXiv:1412.6980. [Published as a conference paper at ICLR 2015].
- [15] G. S. Montavon, S. Bach, A. Binder, W. Samek and K.-R. Müller. "Explaining nonlinear classification decisions with deep Taylor decomposition." *Pattern Recognit.* vol. 65, pp. 211–222, 2017.
- [16] M. Sironi et al. "SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective." *Infection, Genetics and Evolution*, vol. 84, 104384, pp. 1-15, 2020.
- [17] K. Suryamohan et al. "Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2." *Communications Biology*, vol. 4, no. 1, pp. 1–11, 2021.

Analyzing Switch Regions of Human Rab7a and Rab10 by Molecular Dynamics Simulations

Levy Bueno Alves, Sarah Sandy Sun and Silvana Giuliatti*

Department of Genetics, University of São Paulo

Ribeirão Preto, Brazil

e-mail: silvana@fmrp.usp.br

Abstract— Rab7a and Rab10 are small GTPases that regulates cellular processes by alternating between its GDP-bound inactive and the GTP-bound active states. Studies have shown that functional deficiencies in the pathways of these enzymes are implicated in ciliopathies, cancer and neurodegenerative diseases. Thus, the modulation of the activity of these targets may represent an interesting strategy in drug discovery for the treatment of several human diseases. In order to identify potential Rab7a and Rab10 inhibitors, we studied the mobility of the switch1-interswitch-switch2 surface to understand the active “ON” and inactive “OFF” states of these enzymes. We use molecular dynamics simulations to investigate the atomic movements of the switch regions of these enzymes associated with GDP and GTP nucleotides. We found noticeable differences in the local flexibility of switch 1 when these Rab GTPases were bound to GDP. However, the deterministic method used was not able to successfully differentiate the flexibility of switch 2 region. We hypothesized that the flexibility of the switch 1 region can be used as an indicator of in silico studies that search potential competitive inhibitors based on nucleotides against these targets. Furthermore, the present study can be useful for research that involves the description on-to-off process of other target proteins.

Keywords-Small GTPases; structural flexibility; in silico.

I. INTRODUCTION

Rab7a and Rab10 are small monomeric enzymes that belong to the Rab GTPase family. They are responsible for regulating intracellular traffic in various pathways of different cellular sublocations, having roles in the endoplasmic reticulum, trans-Golgi network, endosomes, lysosomes, and primary cilium [1]. Several Rab GTPases are believed to be involved in cellular processes such as biogenesis, transport, and tethering of membrane-bound organelles/vesicles. However, functional dysregulations in the pathways of these enzymes are implicated in a few human diseases, such as ciliopathies, cancer, and neurodegenerative diseases [2]. For example, studies have shown that Rab10 and Rab7a have a relevant role in Alzheimer's disease (AD), where Rab10 helps in the amyloidogenic processing of the amyloid precursor protein (APP) while Rab7a is involved in the hyperphosphorylation of the Tau protein [3] [4]. Furthermore, there are studies that correlate the inhibition of these enzymes with the induction of apoptosis in cancer cells [5]. Experiments show that changes in the expression of these enzymes can activate

signaling pathways for cancer cells growth and survival, leading to cancer progression. Such evidence paves the way for the application of new drug targeting strategies for the treatment of various human diseases. Since Rab7a and Rab10 are associated with several human diseases, the modulation of the activity of these enzymes using small molecules may represent a promising alternative to delay the progression of these diseases, making them potential therapeutic targets.

Rab GTPases regulates cellular processes by alternating the nucleotides guanosine triphosphate (GTP) and Guanosine diphosphate (GDP). When bound to GTP, they interact with a series of effector proteins promoting downstream signaling events. On the other hand, the hydrolysis of GTP results in conformational changes in the G domain of these enzymes, inactivating them [6]. The differences between the conformations of the G domain linked to GDP and GTP suggest that after the hydrolysis of GTP the switch 1 and switch 2 regions show a high degree of flexibility and disorder. In contrast, these regions are stabilized in the active state, which favors their recognition by effector proteins [7]. This occurs due to interactions with the phosphate binding motifs (PMs) of these enzymes, where GTP interacts with PM1-3 while GDP only with PM1 [8] [9].

In 2017, PYLYPENKO and collaborators conducted a study to obtain insights into the functional diversity of Rab GTPases. In this study, the authors analyzed, using in silico tools, 44 representatives of the subfamilies of human Rab GTPases, to obtain information on the primary sequences of these enzymes with partner proteins in the context of binding specificity and provide results functions of their interactions in the cell. In this study, the authors detailed precision the motives of the Rab GTPases that interact with GDP and GTP; however, it was not analyze the “on-to-off” process to understand the modifications conformational patterns of these enzymes when activated and inactivated.

The present study aimed to detail the structural flexibility of the switch regions of Rab7a and Rab10, considering 200 ns molecular dynamics (MD) simulations. In 2020, we performed MD simulations to investigate Rab10's internal movements in its activated and inactivated state. These results showed noticeable differences in local switch I flexibility when Rab10 was associated with GDP [10]. Thus, we hypothesize that the flexibility of the switch1 region can be used as an indicator of in silico studies that aim to identify potential competitive inhibitors against Rab10. In order to verify if the conformational change in the on-to-off process

of Rab10 was not by chance, we extended our study to consider the Rab7a. Our new findings strengthen our hypothesis that the flexibility of the switch I region can be used as an indicator for studies aimed at identifying potential inhibitors of these enzymes.

The rest of this work is organized as follows. Section II describes the methods used in each step of this study. Section III addresses the results and discussion, while Section IV describes the conclusion and next steps of this research. The acknowledgements close the article.

II. METHODS

In this section, the *in silico* approaches used during the development of this work.

A. Molecular docking

The structure of Rab7a (ID: 1T91), Rab10 (ID: 5SZJ) and GDP and GTP nucleotides were obtained from the Protein Data Bank (PDB) [11]. Modeller software v9.23 [12] was used to fill the missing atoms of Rab10. The addition of hydrogen in each structure, considering the protonation state of the atoms at physiological pH, was performed using the Open Babel 3.0.0 software [13]. The Autodock Vina 1.1.2 software [14] was used to docking the nucleotides at the active site of Rab10. The grid box was defined by the mean of the Cartesian coordinates of the nucleotide GTP and phosphoaminophosphonic acid guanylate ester (GNP) co-crystallized in Rab7a and Rab10, respectively. The co-crystallized compounds were submitted to redocking to validate the docking study. The poses of each nucleotide were chosen by means of the lowest binding energy and the highest number of intermolecular bonds. The interactions between the ligands and receptor were calculated using the Maestro 12.3 interface [15].

B. Molecular dynamics

The GROMACS package version 2019.3 [16] was used in the MD simulations of complexes with GDP and GTP. The force field used was CHARMM36 [17]. The ligand parameters were obtained by the CGenFF server [18]. The complexes were centralized in cubic boxes, where the distance between the solute and the edge was 14 Å. The molecules were solvated with TIP3P water molecules and neutralized by adding the appropriate number of Na⁺Cl⁻ ions considering the ionic concentration of 0.15 M. The energy minimization was performed using the steepest descent method with a maximum force of 1000 KJ/mol.nm. After minimization, the systems were equilibrated in two stages: a canonical NVT set (number of particles, volume, and temperature) followed by an isothermal-isobaric NPT set (number of particles, pressure, and temperature). The NVT equilibrium was performed with a constant temperature of 300 K for 500 ps. The NPT equilibrium was performed with a constant pressure of 1 bar and a constant temperature of 300 K for 500 ps. The production step was carried out at 300 K for 200 ns and the trajectories were saved every 10 ps. The tools of the root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg) and

solvent accessible surface area (SASA) were used for the trajectory analysis.

III. RESULTS AND DISCUSSION

The results and discussion of this study are described in subsequent sections.

A. Molecular docking study

The lowest energy values for each test were grouped and their molecular interactions were analyzed. The most promising poses of each ligand are described in Table 1. The comparison between the co-crystallized ligands poses with the docking poses indicated RMSD ≤ 2.00 Å, suggesting that the docking protocol was validated. Thus, GDP and GTP nucleotides were successfully anchored in the active site of Rab7a and Rab10. These complexes showed notable intermolecular interactions. When in interaction with GDP, there are 6 hydrogen bonds with the residues located in PM1, while when interacting with GTP there are more hydrogen bonds.

TABLE I. SCORE BY THE VINA AND THE NUMBER OF INTERACTIONS CALCULATED BY MAESTRO

Complex	Score (Kcal/mol)	Salt Bridge	HBonds	Stacking
Rab7a-GDP	-10.3	3	6	2
Rab7a-GTP	-11.1	3	9	2
Rab10-GDP	-10.7	3	6	2
Rab10-GTP	-11.5	3	10	2

This occurs due to the presence of γ -phosphate in GTP, which interacts with the PM2 and PM3 motifs and can even increase the stability of the switch regions of these enzymes. Furthermore, the binding modes found for each complex are consistent with the interactions found in the Rab7a and Rab10 crystals [11] [12]. This indicates that the docking protocol was able to successfully reproduce the experimentally determined binding mode for the co-crystallized ligands.

B. Molecular dynamics study

To verify whether Rab7a and Rab10 stabilized throughout the MD simulation, the RMSD of the G domain of these enzymes was calculated. Based on our results, the trajectories reached stability after 40 ns of simulation (Fig. 1a). In the case of Rab7a, the backbone RMSD of the complexes associated with GDP and GTP stabilized close to 0.23 and 0.24 nm, respectively. In relation to Rab10, the trajectories stabilized close to values of 0.30 nm for the inactive state and 0.25 nm for the active state. Through analysis of the trajectories, it was possible to observe that the complexes associated with the GTP have more restricted movements, since the trajectories showed predominantly lower peaks than the complexes linked to GDP.

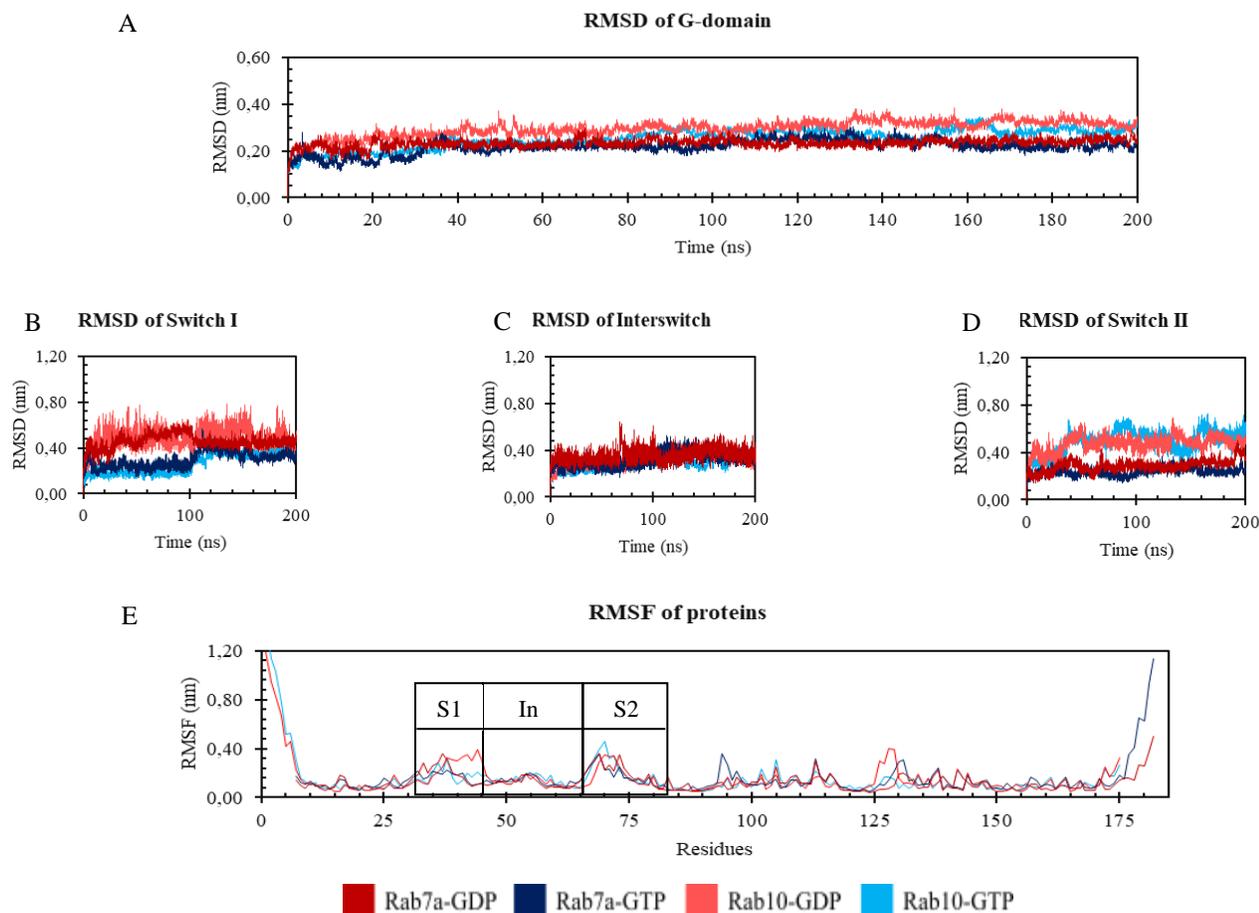


Figure 1. Analysis of the trajectories obtained in the MD simulation. (A) RMSD of G-domain backbone. (B) RMSD of switch I. (C) RMSD of interswitch. (D) RMSD of switch II. (E) RMSF of the amino acids residues. Switch 1 (S1) region is defined by positions 31-44, while interswitch (In) and Switch 2 (S2), 45-65, 66-82, respectively.

Local RMSD analysis was required to assess the inherent flexibility of switch regions. Fig. 1b shows the RMSD of the switch I region, where it is possible to visualize differences in the backbone when the enzymes analyzed were associated with the nucleotide GDP and GTP. Here, systems linked to GDP showed greater fluctuations compared to those linked to GTP. This can be explained because the phosphate portion of GDP contains only α and β -phosphate, causing switch I to suffer greater conformational flexibility due to the existence of smaller intermolecular interactions involving the PM2 and PM3 motifs. On the other hand, the presence of γ phosphate in GTP provides less flexibility in switch I due to the existence of greater intermolecular interactions with these motifs. The interswitch region was chosen as a control because it does not show conformational differences in both the activated and inactivated states of the Rab GTPases (Fig. 1c). However, the switch II region showed no noticeable RMSD differences (Fig. 1d). This result may be associated with the presence of an α -helix in switch II, which can generate motions more restrictive with varying time, as each successive turn of the α -helix is held by adjacent turns by three or four hydrogen bonds, which gives a significant stability in relation to other secondary structures of these

enzymes. Fig. 1e shows the residues that make up the entire extension of the Rab7a and Rab10 enzymes, where it is possible to visualize the regions where the greatest fluctuations occur. These results show subtle differences in the flexibility of switch I, however, in all complexes where Rabs were inactivated, the fluctuation for this region was greater.

Rg trajectories indicate that the G domain of Rab7a and Rab10 entered conformational equilibrium. Fig. 2a shows that the Rg values are constant, suggesting that the folded structures of these enzymes are stable. However, the Rg trajectories of the inactivated systems showed higher values than the activated state. While the difference in these values is subtle, it does indicate that the inactivated state is subject to decompression processes throughout the simulation. Furthermore, SASA analysis was performed to evaluate the molecular surface of these Rabs. Fig. 2b shows a slight increase in SASA for inactivated states, which can be explained due to greater mobility in the switch region I.

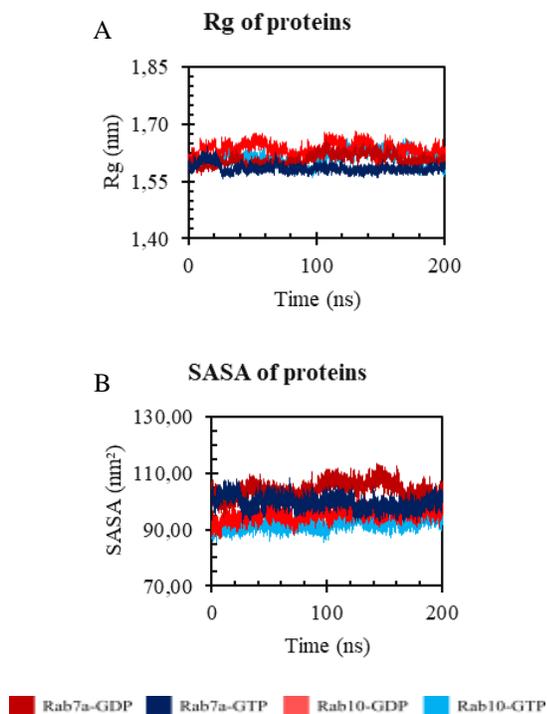


Figure 2. Analysis of compression and SASA. (A) Rg of the G-domain. (B) SASA of the G-domain.

IV. CONCLUSIONS AND FUTURE WORK

In short, the MD simulations used in this study were able to obtain notable differences in the switch 1 region of Rab7a and Rab10, enabling the identification of its active "ON" and inactive "OFF" states. However, the classical mechanics method was unable to predict the disordered movements of the switch 2 region. Our new findings strengthen our hypothesis that the flexibility of the switch I region can be used as an indicator for studies aimed at identifying potential inhibitors of these enzymes. Furthermore, the data discussed here may be useful for research involving the description of the on-to-off process of other proteins. The next steps of this research is to extend the analyzes to evaluate the on-to-off process of 44 representatives of human Rab GTPases, to understand the conformational modifications of these enzymes when activated and inactivated. Thus, it is expected that this study proposal provides useful information about the dynamics of these enzymes, allowing regions switches are used as indicators to select which putative drugs have the potential to inhibit or activate these enzymes.

ACKNOWLEDGMENT

This work was supported by Coordination for the Improvement of Higher Education Personnel (CAPES), National Council for Scientific and Technological

Development (CNPq), and Foundation to support Teaching, Research and Assistance at HCFMRP-USP (FAEPA). The authors acknowledge the Superintendence of Technology and Information at USP for the High-Performance Computing (HPC) Computing Resources.

REFERENCES

- [1] J. Liu, X. Zheng, and X. Wu, "The Rab GTPase in the heart: Pivotal roles in development and disease", *Life Sciences*, vol. 306, no. 120806, pp. 1-8, 2022.
- [2] Y. Homma, S. Hiragi, and M. Fukuda, "RAB family of small GTPases: an updated view on their regulation and functions", *FEBS J.*, vol. 288, pp. 36-55, 2021.
- [3] P. G. Ridge et al., "Linkage, whole genome sequence, and biological data implicate variants in RAB10 in Alzheimer's disease resilience", *Genome Medicine*, vol. 9(1), no. 100, pp. 1-14, 2017.
- [4] L. Rodriguez et al., "Rab7A regulates tau secretion", *Journal of Neurochemistry*, vol. 141(4), pp. 592-605, 2017.
- [5] G. D. P. Krishnan et al., "Rab GTPases: Emerging Oncogenes and Tumor Suppressive Regulators for the Editing of Survival Pathways in Cancer", *Cancers (Basel)*, vol. 12(2), no. 259, pp. 1-20, 2021.
- [6] R. G. Good, M. P. Müller, and Y. Wu, "Mechanisms of action of Rab proteins, key regulators of intracellular vesicular transport", *Biological Chem.*, vol. 398(5-6), pp. 565-575, 2017.
- [7] O. Pylypenko et al., "Rab GTPases and their interacting protein partners: Structural insights into Rab functional diversity", *Small GTPases*, vol. 9(1-2), pp. 22-48, 2018.
- [8] J. Cherfils, and M. Zeghouf, "Regulation of Small GTPases by GEFs, GAPs, and GDIs", *Physiological Reviews*, vol. 93(1), pp. 269-309, 2013.
- [9] L. B. Alves, W. O. Castillo-Ordoñez, S. Giuliatti, "Analyzing Switch Regions of Human Rab10 by Molecular Dynamics Simulations", *Lecture Notes in Computer Science*. Ied., Springer International Publishing, vol. , pp. 215-220, 2020.
- [10] H. M. Berman et al., "The Protein Data Bank", *Nucleic Acids Research*, vol. 28(1), pp. 235-242, 2000.
- [11] A. Sali, and T. L. Blundell, "Comparative protein modelling by satisfaction of spatial restraints", *J. Mol. Biol.*, vol. 234(1), pp. 779-815, 1993.
- [12] N. M. O'Boyle et al., "Open Babel: An open chemical toolbox", *J. Cheminformatics*, vol. 3(1), no. 33, pp. 1-14, 2011.
- [13] M. Trott, and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading", *J. Comput. Chem.*, vol. 31(1), pp. 455-461, 2010.
- [14] Schrödinger Release 2020-3., "Maestro", New York, NY, 2020.
- [15] D. V. D. Spoel et al., "GROMACS: Fast, flexible, and free", *J. Comput. Chem*, vol. 26(1), pp. 1701-1718, 2005.
- [16] J. Huang, and A. D. MacKerell Jr, "CHARMM36 all - atom additive protein force field: Validation based on comparison to NMR data", *J. Comput. Chem.*, vol. 34(1), pp. 2135-2145, 2013.
- [17] K. Vanommeslaeghe, and A. D. MacKerell Jr, "Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing", *J. Chem. Inf. Model.*, vol. 52(12), no. 12, pp. 3144-3154, 2012.

Improving Recovery Engagement for Patients with Substance Use Disorder in the Emergency Department

Kory London, MD
 Department of Emergency Medicine
 Thomas Jefferson University
 Philadelphia, PA
Kory.London@Jefferson.edu

Les Sztandera, PhD
 Kanbar College of Design, Engineering, and Commerce
 Thomas Jefferson University
 Philadelphia, PA
Les.Sztandera@jefferson.edu

Abstract— The overdose crisis our communities are experiencing is a profound and multifactorial challenge to public health. Political headwinds have placed increased scrutiny and support in finding solutions. Emergency departments (ED) play an essential role in the care for patients with substance use disorders (SUD). Peer support, using those with lived experience to assist in recovery, is an emerging tool in hospitals. Combining medical and behavioral health interventions may result in improved outcomes. In a retrospective analysis, an ED based peer support program in five hospitals in Philadelphia engaged 5821 individuals over 25 months. The program has resulted in an increase in direct referral to recovery services from 3% before the program to 20.5%. Peer support is a valuable tool in recovery engagement, further study is required to determine other benefits of peer support and long-term outcomes.

Keywords- Public Health; Substance Use Disorder; Opioid Use Disorder; Emergency Medicine; Addiction; Recovery; Peer Support

I. INTRODUCTION

The United States is experiencing a nearly unprecedented epidemic that transcends social strata and bears no easy answers - the overdose crisis. What grew as a sequela of the dramatic increase in prescription opioids in the early 21st century, this crisis has now entrapped countless individuals in a cycle of addiction [1]. The infiltration of fentanyl and veterinary tranquilizer (xylazine) into the heroin supply and now into the supplies of other drugs has led to a dramatic rise in polysubstance use and the inevitable increase in deaths when individuals mix drugs [2]. We are witnessing an era where overdose deaths greatly exceed those of violent crime [3]. Substance use disorders (SUD) represent a public health emergency that requires novel solutions to these exceptional problems.

Emergency Departments are legally required to see all patients who arrive for evaluation [4]. The risk of overdose, on top of other medical complications of SUD, including skin, spinal and heart valve infections as well as complex wounds [5], lead many patients to visit emergency departments for care [6]. Emergency departments are manifestly focused on discovering and treating conditions that threaten the immediate health of their patients, leaving advocacy for chronic conditions, such as SUD, an afterthought. National data show fewer than 2% of patients with SUD received treatment within the past year [7].

Peer support is an emerging trend in recovery services [8]. Utilizing individuals in mature phases of recovery to

engage and advocate for recovery is important for a number of reasons. Primarily, individuals who have not experienced SUD often have difficulty understanding and empathizing with those currently suffering. Secondly, even if engaged by an empathetic and caring provider, many individuals have experienced traumatic healthcare experiences or believe recovery to be impossible, limiting their engagement. Peers have immediate credibility in these situations and can act as experiential interpreters for these patients and their caregivers. Lastly, navigating the recovery ecosystem is a time consuming and often byzantine journey, one that does not mesh well with the fast paced, constantly task switching environment of the ED. Peers can be tasked with this as their primary work, freeing others to provide care while their recovery is still addressed.

Given the above, it is unsurprising peers are entering the hospital environment [9]. Peers have been shown to improve uptake of buprenorphine, a medication for opioid use disorder (MOUD) [10]. They also have been shown to improve rates of discharge with naloxone, a medication that can reverse the effects of overdose [11]. They are relevant to visits both related to overdose and other reasons for hospitalization [12]. They are also able to address other social determinants of health (SDOH), another emerging and related topic in social medicine [13]. Their overall impact on referral to recovery services, especially for non-opioid use disorders, is still being defined.

This study aims to evaluate a public health initiative placing peer support specialists in a series of urban emergency departments. We describe the demographics of patients who are engaged by peers, their primary drugs of choice, and their recovery outcomes. Additional data discussed includes rates of referral to SDOH resources. Lastly, we discuss the impact of CRS in the ED and wider hospital-based healthcare ecosystem.

II. METHODS

This represents the collected data from the first 25 months of an emerging public health intervention. Left unstated are the dedicated contributions of many in the development, deployment and continued function of this program.

A. Study Population

The program represents a collaborative partnership between Jefferson Health and the City of Philadelphia Department of Health and began in December 2020. Five emergency departments were included: two urban academic hospitals, two urban community hospitals and one suburban community hospital.

Certified recovery specialist (CRS) is the term given by the Commonwealth of Pennsylvania for these peers. CRS are individuals in long term recovery, who partake in didactic and practical classes to improve their skills in engaging patients about recovery services, supporting individuals in early recovery and navigating the recovery system. CRS placed in the ED are able to meet with patients who are admitted to the hospital with serious medical conditions as well as those being discharged after their ED visit. Importantly, while the focus of the program was engaging patients with opioid use disorders, CRS are able to assist patients regardless of primary substance. They work collaboratively with providers, nurses and social workers to help improve patient care.

Referrals to the CRS are low barrier, akin to pastoral care and social work consultations and occur both through the electronic health record and via role cellphone. They can be provided by any staff member who interacts with a patient who may benefit from a CRS visit, as long as the patient is able to consent to meeting with one. Patients both in the ED and on the inpatient units were engaged by CRS, providing a robust and diverse patient cohort.

B. Study Design

This project is a retrospective analysis of a cohort of patients seen in the Jefferson Health emergency departments in the greater Philadelphia, PA area. Data was collected via two separate sources. The electronic health records of the patient, EPIC (EPIC Systems, Madison, WI) as well as a separate encrypted record system that organized data requested by the city, LAURIS (LAURIS Online, Roanoke, VA). CRS met with patients, speaking on a variety of topics, including the challenges of being in the hospital, the different options for MOUD and other recovery support services including group and solo therapy and family outreach. They are also able to assist with other SDOH, such as lack of identification documents, housing instability, food insecurity and employment services.

Data is collected by the CRS if the patient verbally consents to the encounter. Patient demographics (age, gender, race), primary drugs of choice, results of their recovery engagement and referrals for SDOH services are the primary variables measured. Data included in this study track from the onset of the program in December 2020 through January 2023. This study has been reviewed by the relevant IRB and deemed exempt.

III. RESULTS

461910 ED visits occurred from December 2020 to January 2023 in the five hospital sites. Of these, 23766 received behavioral health diagnoses (5.1%). During the study period, a total of 5821 patients were engaged at least once by a CRS (24.5%) (Table 1). The mean age of the patients was 42 years (SD = 11.4) and 4015 (69.0%) identified as male (30.8% female, 0.2% transgender or other). 3379 (58.0%) reported their primary drug of choice were opioids (27.0% alcohol, 10% stimulants, 5.0% others).

TABLE I. PATIENT DEMOGRAPHICS

Patient Demographics	
Age (Mean)	42
Engagements	5,821
Hospital	
TJUH	2583 (44.0)
Frankford	1824 (31.0)
Torresdale	793 (14.0)
Methodist	451 (8.0)
Bucks	170 (3.0)
Gender	
Male	4015 (69.0)
Female	1796 (30.8)
Transgender	10 (.2)
Drug of Choice	
Opioids	3379 (58.0)
Alcohol	1567 (27.0)
Stimulants	600 (10.0)
Other	275 (5.0)
Patient Disposition	
Referred to Tx	1,195 (20.5)
Incomplete	993 (17.0)
Refused Tx	3,633 (62.5)
SDOH Referrals	1,486 (25.0)

Internal data from 2019 showed that 3% of identified patients with SUD were referred for recovery services from the hospital emergency department. During the study period, 1195 patients were referred directly to recovery services (20.5%), 3633 patients refused referral (62.5%) to recovery

services, and 993 were interested but were without referral at the time of their hospital disposition (17.0%). An additional 1495 patients received referrals to services that address SDOH in addition to their recovery conversations. Of the 5821 patients engaged by a CRS, 3376, a majority were seen in an academic hospital emergency department (58.0%). The largest cohort seen in a community hospital was at the hospital located closest to the epicenter of the overdose crisis in Philadelphia (31.0%) and the smallest cohort engaged was present at the suburban community hospital (3.0%).

IV. DISCUSSION

Many patients with chaotic substance use and SUD actively avoid hospitals until they have grave health consequences. Matching medical and behavioral treatment in a multidisciplinary fashion is becoming more common as hospitals try to treat patients holistically, rather than in a problem-based fashion. By engaging patients when they are experiencing severe sequelae of their use, health systems may better be able to engage patients. A framework of this strategy is shown in Figure 1. In this retrospective, multicenter cohort evaluation, a large number of patients receiving ‘acute unplanned care’ were able to get concomitant medical and behavioral health services in the emergency department and hospital setting. The cohort’s demographics largely matches national studies of patients with substance use disorder: white, male and young-middle aged [14]. It is important to note that many minority populations receive disparate medical and behavioral health care [15]. Assuring that these programs are able to engage and assist all eligible patients is imperative to reducing these disparities.

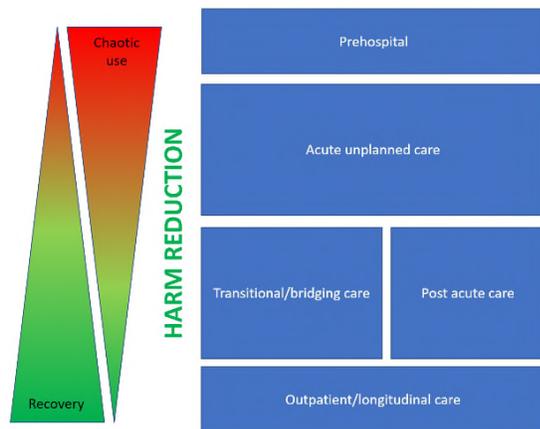


Figure I. Medical/Behavioral Engagement Framework

Much of the focus on SUD related care comes from the increased awareness and scrutiny that surrounds the care of patients with opioid use disorders. The reasons for this are numerous, but include the excess mortality related to OUD as compared other substances, the insidious and questionable

practices of pharmaceutical companies that led to opioids being prescribed for acute and chronic pain and the infiltration of fentanyl and other synthetic opioids into other drug supplies. While the program was officially developed as a response to the opioid overdose crisis, it is also important to note that almost half of the patients engaged by the CRS had a non-opioid primary drug of choice.

Successfully connecting patients to recovery services is a well-documented challenge. As previously stated, less than 2% of patients nationally with SUD receive any recovery services and previous programs prior to this intervention referred a total of 3% of identified patients. By utilizing CRS, referrals to recovery services increased almost seven-fold. This represents a crucial success, especially sustaining that level of improvement over two full years. Saying that, nearly two thirds of patients refused recovery services and an almost equal number to those successfully referred were unable to be connected, due to lack of availability or other logistical barriers. While CRS add clear value, there are many patients who an alternate approach may be appropriate [16].

It is sadly common that patients with multiple SDOH issues frequently visit the ED, regardless of concomitant SUD [17]. Whether afflicted with homelessness, food insecurity or lack of access to primary care, many challenges associated with SUD also bring patients to the ED. The ability to engage patients with other social services acts as a force multiplier for the CRS. Addressing addiction is a critical feature of recovery but must be met with a holistic system that addresses the cycles of trauma and other barriers to long term recovery. By actively addressing these SDOH, patient behavior is reinforced and the process to get appropriate individuals integrated into the public health system is supported.

There are several limitations to this study. First, the data reported is retrospective, limiting the variables that can be studied. The trial is also uncontrolled, meaning it is possible that the impact witnessed was related to another, unrelated aspect of our care. Third, the city’s system mandated patients provide a single ‘primary drug of choice’, despite many of our patients explicitly being polysubstance users. While it appears the majority of our patients used opioids primarily, it is possible they are also using other drugs. Recognition of polysubstance use is increasingly important and impacts both the withdrawal symptoms the patients encounter and the complexity of their recovery support. Lastly, the definition of referral was made by the city and relates to services the patients goes to directly from the hospital. A fair number of the patients in the ‘incomplete’ row received recovery services subsequent to their hospitalization as a result of the CRS engagement. It is highly likely the cited figure underrepresents the number of patient referrals.

Despite the limitations, we believe the study provided much needed insights to all of the stakeholders involved. By actively encouraging involvement in the implementation

strategies, we ensured widespread buy-in and increased the odds of making an impact on the community health. The support of hospital leadership was essential for securing funding and resources to implement desired strategies. Externally, engaging the local / city health administration, and reinforcing and strengthening the relationship was essential as hospitals moved from the assessment phase to developing and implementing strategies to address identified community health priorities.

The Collective Impact Framework indicates that no single entity or department alone can address the society's most complex and challenging problems [18]. The health needs identified in the conducted study were indeed the result of complex social, economic, as well as environmental factors, making Collective Impact Framework (Fig. 1) an appropriate model to apply. Widespread collaboration among community stakeholders around shared health challenges could reinforce positive changes in the community. While these engagements and referrals play an important initial role in recovery, it will require a more longitudinal study to determine the ultimate outcomes of an intervention like the one described.

V. CONCLUSION

Peer support is an emerging tool to improve patient engagement for recovery services in a hospital setting. The challenges involved in connecting patients to services from the hospital are myriad for patients and healthcare workers. Peers are able to engage with a variety of individuals with varied SUD and provide both in-hospital recovery support as well as referrals to a variety of services. In this retrospective cohort study, there are positive signals indicating that peers improve connection to recovery and social services. By matching medical and behavioral therapies, patients' needs will be better serviced. Future studies should evaluate harm reduction strategies for CRS in patients who refuse recovery support, patient attitudes towards peer support, and barriers to recovery referral.

ACKNOWLEDGMENT

KSL would like to acknowledge Dr. Rebecca Jaffe for development of the medical/behavioral engagement framework.

REFERENCES

- [1] Kolodny, Andrew, David T. Courtwright, Catherine S. Hwang, Peter Kreiner, John L. Eadie, Thomas W. Clark, and G. Caleb Alexander. "The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction." *Annual review of public health*, 559-574, 2015.
- [2] Friedman, Joseph, Fernando Montero, Phillippe Bourgois, Rafik Wahbi, Daniel Dye, David Goodman-Meza, and Chelsea Shover. "Xylazine spreads across the US: a growing component of the increasingly synthetic and polysubstance overdose crisis." *Drug and alcohol dependence*, 233, 109380, 2022.
- [3] Centers for Disease Control. *Drug Overdose Deaths in the U.S. Top 100,000 Annually*. [Online] Available from https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm
- [4] Bitterman, Robert A. "EMTALA and the ethical delivery of hospital emergency services." *Emergency Medicine Clinics* 24, no. 3, 557-577, 2006.
- [5] Malayala, Srikrishna V., Bhavani Nagendra Papudesi, Raymond Bobb, and Aliya Wimbush. "Xylazine-Induced Skin Ulcers in a Person Who Injects Drugs in Philadelphia, Pennsylvania, USA." *Cureus* 14, no. 8, 2022.
- [6] Zhang, Xingyu, Ningyuan Wang, Fengsu Hou, Yaseen Ali, Aaron Dora-Laskey, Chin Hwa Dahlem, and Sean Esteban McCabe. "Emergency department visits by patients with substance use disorder in the United States." *Western Journal of Emergency Medicine* 22, no. 5, 1076, 2021.
- [7] SAMHSA. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health*. Available from: <https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm>
- [8] McGuire, Alan B., Kristen Gilmore Powell, Peter C. Treitler, Karla D. Wagner, Krysti P. Smith, Nina Cooperman, Lisa Robinson, Jessica Carter, Bradley Ray, and Dennis P. Watson. "Emergency department-based peer support for opioid use disorder: Emergent functions and forms." *Journal of substance abuse treatment* 108, 82-87, 2020.
- [9] Shalaby, Reham A. Hameed, and Vincent IO Agyapong. "Peer support in mental health: literature review." *JMIR mental health* 7, no. 6, e15572, 2020.
- [10] McGuire, Alan B., Kristen Gilmore Powell, Peter C. Treitler, Karla D. Wagner, Krysti P. Smith, Nina Cooperman, Lisa Robinson, Jessica Carter, Bradley Ray, and Dennis P. Watson. "Emergency department-based peer support for opioid use disorder: Emergent functions and forms." *Journal of substance abuse treatment* 108, 82-87, 2020.
- [11] Ramdin, Christine, Marshall Guo, Scott Fabricant, Cynthia Santos, and Lewis Nelson. "The impact of a peer-navigator program on naloxone distribution and buprenorphine utilization in the emergency department." *Substance Use & Misuse* 57, no. 4, 581-587, 2022.
- [12] Ashford, Robert D., Matthew Meeks, Brenda Curtis, and Austin M. Brown. "Utilization of peer-based substance use disorder and recovery interventions in rural emergency departments: Patient characteristics and exploratory analysis." *Journal of rural mental health* 43, no. 1, 17, 2019.
- [13] Anderson, Erik S., Suzanne Lippert, Jennifer Newberry, Edward Bernstein, Harrison J. Alter, and Nancy E. Wang. "Addressing social determinants of health from the emergency department through social emergency medicine." *Western Journal of Emergency Medicine* 17, no. 4, 487, 2016.
- [14] SAMHSA. *Key Substance Use and Mental Health Indicators in the United States: Results from the 2019 National Survey on Drug Use and Health*. Available from: <https://www.samhsa.gov/data/sites/default/files/reports/rpt29393/2019NSDUHFFRPDFWHTML/2019NSDUHFFR090120.htm>
- [15] Sanchez, Katherine, Rick Ybarra, Teresa Chapa, and Octavio N. Martinez. "Eliminating behavioral health disparities and improving outcomes for racial and ethnic minority populations." *Psychiatric Services* 67, no. 1, 13-15, 2016.
- [16] Taylor, R. Morgan, and Cynthia S. Minkovitz. "Warm handoffs for improving client receipt of services: A systematic review." *Maternal and child health journal* 25, 528-541, 2021.
- [17] Doran, Kelly M., Nathan M. Kunzler, Tod Mijanovich, Samantha W. Lang, Ada Rubin, Paul A. Testa, and Donna Shelley. "Homelessness and other social determinants of health among emergency department patients." *Journal of Social Distress and the Homeless* 25, no. 2, 71-77, 2016.

[18] Hanleybrown, F., Kania, J., and Kramer, M. (2012).
Channeling change: Making collective impact work.
Available

from: <http://icisd.org/cms/lib/MI01928326/Centricity/Domain/218/Making%20Collective%20Impact%20Work%20Stanford%202012.pdf>

Examining the Relationship between COVID-19 Mobility and Eviction Rates in Philadelphia

Regina Ruane

The Wharton School
The University of Pennsylvania
Philadelphia, PA, USA
e-mail: ruanej@upenn.edu

Les Sztandera

Kanbar College of Design, Engineering, and Commerce
Jefferson University
Philadelphia, PA, USA
e-mail: Les.Sztandera@jefferson.edu

Abstract—The COVID-19 pandemic has had a significant impact on public health, the economy, and social norms. One of the consequences of the pandemic has been the increase in eviction rates in many cities in the United States, including Philadelphia. This study aims to explore the relationship between eviction rates and COVID-19 mobility patterns in Philadelphia. We analyzed eviction data from the city of Philadelphia and mobility data from Google's COVID-19 Community Mobility Reports. Our findings suggest that there is a statistically significant relationship between eviction rates and mobility patterns. Specifically, we found that areas with high eviction rates also had a higher level of mobility, which could potentially increase the spread of the virus. Our results highlight the importance of considering the impact of socioeconomic factors on the transmission of COVID-19.

Keywords—COVID-19, eviction, mobility.

I. INTRODUCTION

The COVID-19 pandemic has affected people across the globe, causing millions of deaths and economic instability. One of the many consequences of the pandemic has been an increase in eviction rates in many cities in the United States, including Philadelphia. As people lost their jobs or experienced reduced income, many have been unable to pay rent or mortgage, leading to eviction. Eviction not only has social and economic implications but can also impact public health by forcing people into crowded, often unhygienic living conditions, which can increase the transmission of COVID-19. The COVID-19 pandemic has brought about unprecedented mobility restrictions to prevent the spread of the virus. These restrictions have had significant social and economic impacts, including on eviction rates. This paper examines the impact of COVID-19 mobility restrictions on eviction rates in Philadelphia, Pennsylvania. Using eviction data from the Eviction Lab and

mobility data from Google's COVID-19 Community Mobility Reports, we conduct a comparative analysis of eviction rates before and after the implementation of mobility restrictions in Philadelphia. Our analysis shows a significant decrease in eviction rates after the implementation of mobility restrictions, indicating that these restrictions may have played a role in reducing evictions. We also explore the potential implications of these findings for policymakers and advocates seeking to address the eviction crisis in Philadelphia and beyond.

The COVID-19 pandemic has exposed and exacerbated existing socioeconomic and health disparities, including disparities in health and well-being. Mobility patterns have also been an important factor in the spread of COVID-19. Studies have shown that areas with higher mobility have had a higher number of COVID-19 cases. Understanding the relationship between eviction rates and mobility patterns can provide insights into how socioeconomic factors can impact the transmission of COVID-19.

Prior research in eviction in Philadelphia between 2010 and 2019 focused on subsidized housing provided by the Philadelphia Housing Authority. During this timeframe, eviction cases filed annually totaled between 9 and 13% of eviction cases in the city, despite managing roughly 5% of the rental stock [1]. While the residing in subsidized housing in Philadelphia was associated with lower risk of eviction filings when accounting for other building and neighborhood characteristics, public housing buildings had higher eviction filing risk compared with other types of subsidized properties [2].

The COVID-19 pandemic has disrupted life as we know it, with governments around the world implementing unprecedented measures to limit the spread of the virus. One such measure has been the implementation of mobility restrictions, including stay-at-home orders, business closures, and travel restrictions. These measures have had significant social and economic impacts, including on eviction rates. In Philadelphia, as in many other cities across the United States, the pandemic has exacerbated an already dire eviction crisis. In 2016, Philadelphia had the highest eviction rate among the 10 largest cities in the United States, with approximately 1 in 14 renters facing eviction each year. Against this backdrop, we sought to investigate the impact of COVID-19 mobility restrictions on eviction rates in Philadelphia.

Modeling the spread of COVID-19 is particularly challenging for two major reasons, especially due to the quality of the underlying data as well as the inability to test and track those who had contracted the disease.

II. METHODS

We collected eviction data from the city of Philadelphia for the period between March 2020 and December 2021. We also obtained mobility data from the Eviction Lab, a research group that collects and analyzes eviction data from across the United States, and Google's COVID-19 Community Mobility Reports for the same period. The mobility data included information on the number of visits to different categories of places, such as retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. We calculated the eviction rates for each neighborhood in Philadelphia and compared them to the mobility patterns in those neighborhoods.

Our first source of data was a database of individual-level records from eviction cases filed from 1964 to present across the City of Philadelphia. The records were provided by the City of Philadelphia and contained case-specific

information, including the court in which the case was filed, court-assigned case number, dates associated with case actions, such as the case filing date, plaintiff (landlords) name(s), defendant (tenant) name(s) and addresses, and an indicator of whether the defendant represented an individual or business. Plaintiff names recorded the party who filed the case.

Case filings were represented by the court identifier and case number. Many cases were represented by multiple individual-level records associated with different defendants or actions. We aggregated filings annually by the earliest date on a record associated with a case. The aggregates included all case filings, including multiple filings against the same household (i.e., serial filings). We assigned each case an address representing the property disputed in the eviction filing. Addresses were cleaned and geocoded. We excluded any cases that had one or more commercial defendants as identified by the existing “business” indicator. We also removed cases that duplicated the same dates, plaintiff names, and tenant addresses across cases.

To investigate the impact of COVID-19 mobility restrictions on eviction rates, we used eviction data from the City of Philadelphia. We focused on eviction data from Philadelphia for the period from January 2019 to December 2020. We also used mobility data from Google's COVID-19 Community Mobility Reports, which provide anonymized data on mobility trends in different categories of places, such as retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. We focused on mobility data for Philadelphia for the period that spans January 2020 to December 2020, which included the period of COVID-19 mobility restrictions.

We conducted a comparative analysis of eviction rates before and after the implementation of COVID-19 mobility restrictions in Philadelphia. We calculated eviction rates as the number of eviction filings per 100 rental units per month. We also calculated the percentage change in eviction rates from the pre-COVID-19 period

(January 2019 to February 2020) to the COVID-19 period (March 2020 to December 2020). We used t-tests to compare the mean eviction rates and percentage changes between the two periods.

To investigate the impact of COVID-19 mobility restrictions on eviction rates, we used eviction data from the Eviction Lab, a research group that collects and analyzes eviction data from across the United States. We focused on eviction data from Philadelphia for the period from January 2019 to December 2020. We also used mobility data from Google's COVID-19 Community Mobility Reports, which provide anonymized data on mobility trends in different categories of places, such as retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas. We focused on mobility data for Philadelphia for the period from January 2020 to December 2020, which included the period of COVID-19 mobility restrictions.

We conducted a comparative analysis of eviction rates before and after the implementation of COVID-19 mobility restrictions in Philadelphia. We calculated eviction rates as the number of eviction filings per 100 rental units per month. We also calculated the percentage change in eviction rates from the pre-COVID-19 period (January 2019 to February 2020) to the COVID-19 period (March 2020 to December 2020). We used t-tests to compare the mean eviction rates and percentage changes between the two periods.

III. RESULTS

A. Analysis

Our analysis revealed that areas with high eviction rates also had a higher level of mobility, particularly in places such as retail and recreation, grocery and pharmacy, and parks. Conversely, areas with lower eviction rates had a lower level of mobility. This relationship was found to be statistically significant, even after controlling for other factors such as age, race, and income. These results suggest that the eviction rates and mobility patterns are closely linked, and areas with high eviction rates may experience increased

transmission of COVID-19 due to higher mobility.

Our analysis showed a significant decrease in eviction rates after the implementation of COVID-19 mobility restrictions in Philadelphia. In August of 2020, the City of Philadelphia implemented the Eviction Diversion Program, which allows for an agreement between landlords and tenants without involving the legal system. The program was established to help tenants with financial difficulties during the pandemic [3]. Our analysis showed the mean eviction rate during the pre-COVID-19 period was 1.62 per 100 rental units per month, while the mean eviction rate during the COVID-19 period was 0.96 per 100 rental units per month. This represents a 41.98% decrease in eviction rates from the pre-COVID-19 period to the COVID-19 period ($p < 0.001$). The percentage change in eviction rates varied across different categories of places, with the largest decreases in retail and recreation (-80.23%), transit stations (-72.27%), and workplaces (-54.06%) ($p < 0.001$ for all).

Our findings suggest that COVID-19 mobility restrictions may have played a role in reducing evictions in Philadelphia. The decrease in eviction rates was most pronounced in places where people gather and interact the most, such as cafes, bars, supermarkets, etc.

IV. CONCLUSION

Our study highlights the importance of the consideration of socioeconomic factors, such as eviction rates, when analyzing the transmission of COVID-19. Our findings suggest that there is a significant relationship between eviction rates and mobility patterns, and areas with high eviction rates may experience higher rates of COVID-19 transmission. Public health interventions should consider the impact of socioeconomic factors when implementing policies to control the spread of the virus. Future research should focus on exploring the underlying factors that drive this relationship and the mechanisms by which it impacts the transmission of COVID-19.

REFERENCES

- [1] I. Goldstein, E. Dowdall, C. Weidig, J. Simmons, and B. Carney. "Evictions in Philadelphia: A data & policy update", 2019, *The Reinvestment Fund*. Retrieved from https://www.reinvestment.com/wp-content/uploads/2019/10/ReinvestmentFund_PHL-Evictions-Brief-Oct-2019.pdf.
- [2] G. Preston, and V. J. Reina. "Sheltered from eviction? A framework for understanding the relationship between subsidized housing programs and eviction." *Housing Policy Debate* 31, no. 3-5, 2021, pp. 785-817.
- [3] A. Heinrichs, and M. Treskon, 2023, "Diverting Eviction-Related Cases Away from Courts".

Altering the Behaviour of the *Catostylus mosaicus* Jellyfish using Electromagnetic Fields

Alberto Ivars¹, Francisco Javier Diaz¹, Lorena Parra¹, Eva S. Fonfría², Cesar Bordehore^{2,3}, Sandra Sendra¹, Jaime Lloret^{1*}

¹Instituto de Investigación para la Gestión Integrada de Zonas Costeras, Universitat Politècnica de València, 46730 Grau de Gandia, Spain

²Instituto Multidisciplinar para el Estudio del Medio Ramon Margalef, Universidad de Alicante, 03690 Sant Vicent del Raspeig, Spain

³Departamento de Ecología, Universidad de Alicante, 03690 Sant Vicent del Raspeig, Spain

Email: aivapal@epsg.upv.es, fjdiabla@upv.edu.es, loparbo@doctor.upv.es, eva.fonfria@ua.es, cesar.bordehore@ua.es, sansenco@upv.es, jlloret@dcom.upv.es

Abstract— Jellyfish bloom implies an economic issue directly affecting tourism, fishing, aquaculture, and oil plants. Although many methods have been used to repel jellyfish arrivals to our coasts, such as nets or water currents, none seems to be a permanent solution. In this paper, we test the effect of applying a medium-frequency Electromagnetic (EM) field on jellyfish behaviour, specifically, their movement. A coil inside a jellyfish aquarium is used to generate the EM field. A current generator powers the coil. As an indicator of jellyfish movement, jellyfish pulsations are counted in the presence or absence of the EM field. Selected jellyfish species (*Catostylus mosaicus*) present two types of colouration, blue and brown. According to the results, blue jellyfish showed different behaviour than brown jellyfish. Blue jellyfish have about 50 pulsations without an EM field, decreasing slightly in the presence of an EM field. On the other hand, brown jellyfish pulses about 40 times without an EM field, dropping to values below 20 when the EM field is applied. As suggested by the obtained data, we can propose different systems to prevent jellyfish from getting close to the shore, based on the use of coils to generate EM fields and bottom water current.

Keywords- Medusozoa; Medium-frequency, Marine animals; Motion, Anti-jellyfish barrier

I. INTRODUCTION

The growing bloom of jellyfish in some areas and other gelatinous organisms can provoke multiple economic problems [1]. Moreover, some jellyfish stings can be cause injuries and even deaths [2]. Regarding the economic impact, we must highlight that the impact of jellyfish blooms in tourist destinations, which could reduce the number of visitors due to the fear of jellyfish stings. Jellyfish also generate problems in other sectors such as fishing, where they clog the nets; aquaculture, in which jellyfish can kill fry; and desalination and refrigeration of different industries (power and oil plants), where they obstruct the pipes [3]. Overfishing, water eutrophication, modification of coastal areas and habitats, and climate change are some factors to which these increasing blooms of jellyfish are attributed [4].

Different systems are currently being developed and used to avoid the interaction between humans and jellyfish.

These systems consist of meshes, nets or water currents that prevent the passage of jellyfish to the shore. Nevertheless, these methods are costly, and data obtained in different tests do not confirm their effectiveness [5]. It has been shown that with a 25 mm mesh, the presence of jellyfish is significantly reduced. The number of jellyfish would be further reduced with a smaller mesh. Nonetheless, it would increase the number of detachments of the seabed vegetation, the entanglement of the fauna in the area, and the accumulation of rafts of marine debris.

The effectiveness of these methods might depend on the area in which they are deployed. They were proven as a successful system for preventing jellyfish in areas where the energy of the waves, wind, and tides was low. In areas with higher hydrodynamic conditions, modifications should be made to increase the success rate, which leads to a higher cost [6]. Considering the normal hydrodynamic conditions on our shores, the shallow water current pushes the jellyfish to the shores. If the motion of jellyfish can be altered, the bottom water currents will return the jellyfish offshore.

Electromagnetic (EM) fields have been used for behaviour change experimentation in animals and plants. In the case of animals, low-frequency EM fields have come to modify the behaviour. On the one hand, EM fields were used to see if they affected the orientation of birds, concluding that individuals exposed to EM waves tended to become disoriented [7]. On the other hand, the EM fields were used with seeds of two plant species. The selected species were *Rhododendron smirnowii* and *Morus nigra*. In the case of the first species, seed germination increased by up to 70%. However, in the case of the second species, the germination percentage decreased by 24%. Therefore, according to these two examples, it can be seen that EM fields can affect the behaviour of living beings.

The aim of the paper is to verify if medium-frequency EM fields can alter the behaviour of jellyfish, particularly their motion (pulsation rate), to explore possible solutions to the problems they generate on the shore. These EM fields are expected to modify the motion of *Catostylus mosaicus*, a common jellyfish in Australia. The EM waves are expected to reduce their pulse rate and remain motionless on the bottom, avoiding advection by currents. The conducted tests

consist of exposing the jellyfish to an EM field generated with a coil. The exposure time to the EM field was set at five minutes. The pulse rate is used as an indicator of the jellyfish's motion. Twelve individuals of *C. mosaicus* were selected. Half of the jellyfish were bluish, while the other half was brown. It should be added that the size of the selected individuals was between 2.78 and 4.63 cm.

The rest of the paper is structured as follows; Section 2 outlines the related work. Section 3 details the test bench. The results are discussed in Section 4. Finally, Section 5 summarises the conclusion and future work.

II. RELATED WORK

In this section, we will summarise the current proposals for jellyfish barriers and the existing studies about the use of EM fields on animals.

A. Jellyfish barriers

In [6] Vasslides et al. test the efficiency of anti-jellyfish barriers. Their results indicate that the captured bay nettles in the area where the any-jellyfish barrier net was installed decreased by 28 to 67 %. In addition, the jellyfish captured in the area protected by the net was significantly smaller than in the adjacent areas. Nevertheless, this type of barrier's impact on fauna and flora and its relatively low effectiveness precludes its general use. The successfulness of air bubbles curtains in preventing the passing of jellyfish into aquacultural facilities was tested in 2021 by Haberlin et al. [9]. Their results were not entirely satisfactory. While high airflow reduces the number of jellyfish, low airflow does not affect the amount of jellyfish that passes the curtain. Moreover, the hydrodynamics link to waves increases the number of jellyfish that passes through the curtain.

In a recent survey, N. Killi et al. [10] concluded that early warning systems and barriers are being used in some regions of Spain and France. These barriers are designed for *Pelagia noctiluca* blooms. Another survey by T. A. Morsy et al. [11] indicates that both stinger nets and stinger suits have a limited impact on jellyfish avoidance. While barriers do not affect tiny jellyfish, the suits' performance has not been demonstrated in trials.

B. EM field effects on animals

In this subsection, we detail the existing reported effects of EM fields on animals. Several papers have been published in which the effect of EM fields on different animals is evaluated. There is an increasing tendency for this sort of study in marine animals due to the high number of high-power cables placed in the oceans. Although much literature can be found on this topic, few papers presented a proper comparison of animals' mobility of performance when exposed to the EM fields.

In 2020 Z. L. Hutchison et al. evaluated the behaviour of *Leucoraja erinacea* and *Homarus americanus* (little skate and American lobster) when exposed to anthropogenic EM fields [12]. The selected EM field was the emissions of subsea high voltage cables for domestic electricity supply with a high voltage direct current. They analyse the impact of the EM field on animal mobility. The indicators used were

the total distance travelled, the mean speed, and the proportion of significant turns, among others. All factors were affected by the EM field being considered a difference statistically significant for little skate. The EM field did not affect the lobster's total travelled distance and speed of movement. Similar tests were conducted with european lobster (*Homarmus Gammarus*) by R. Taormina et al. in the same year [13]. In these tests, the authors exposed juvenile individuals to magnetic fields generated by AC/DC submarine power cables. The mobility indicators used were the activity ratio, the mean velocity, and the distance travelled. Their results indicate that no differences in European lobster mobility were detected.

The impact of EM fields on fish was studied by D. P. Fey et al. in 2019 [14]. The authors exposed the eggs of rainbow trout to an EM field for 36 days. Their objective was to detect if the EM field might affect fish development in the early stage. The measured parameters included time to hatching, mortality, and growth rate, among others. Their results indicated no variation between the control and exposed fish. Even there are few differences. Those differences were not statistically significant. The effect of the EM field in simple animals, such as on placozoa, a basal form of a marine multicellular organism, was evaluated in 2020 by A. V. Kuznetsov et al. [15]. They used a square pulse with shallow frequency. As a mobility indicator, the fraction of immobilised individuals of *Trichoplax sp.* was considered. Results indicate that the higher the voltage, the greater the number of immobilised individuals. The exposition time also impacts the number of immobilised individuals greater after 1000 ms than after 1 ms.

Numerous surveys have been published in recent years with mixed results. The main conclusion is that no apparent EM field effect has been demonstrated on marine animals. The main problem of existing research is the differences among studied EM fields and animals. In [12] and [13], scientists used anthropogenic EM fields generated by high-voltage cables with frequencies that range from 1 Hz to 2.5 kHz [12]. There is no information about the frequency in [13]. Meanwhile, in [15], a specific EM field was generated for the experiment with an Arduino Uno microprocessor that generates rectangular pulses with a frequency below 2 kHz. As far as we know, no publication has been found in which the effect of EM fields on jellyfish is studied.

III. TEST BENCH

In this section, the complete test bench is detailed. Firstly, the type of coil used and the current generator have been described to quantify each jellyfish pulsation. Next, the jellyfish used in the tests are described. Later, the followed methodology in the conduct tests is explained. Finally, the used indicator, the pulse rate, is defined.

A. EM field characterisation

To generate the EM field, a coil is used. According to Ampere Law, Eq. (1), it is possible to estimate the generated EM field. The magnetic flux density (B) inside a coil is proportional to the magnetic permeability of the water (μ_0), and the characteristics of the coil, including the number of

spires (N) and the length of the coil (l), and current (I_c). In our scenario, de magnetic flux density in the centre of the coil is 10.3 mT.

$$B = \mu_0 I_c N/l \tag{1}$$

B. Equipment

In order to keep the specimen during the tests, we used a cubic tank. The aquarium was filled with 18 L of seawater, with an average temperature of 21.6 ± 1.1 °C and 31.8 ± 0.42 ppm of salinity, from the aquarium where the jellyfish were maintained; see Figure 1.

As explained before, the coil will be powered with a current generator AFG1022 from Tektronix [16]. The electronic circuit to power the coil includes a resistance of 100 KΩ with a 5% tolerance on the positive side of the copper coil. The generated signal to power the coil was a sinus signal of 3.3 Vpp.

C. Jellyfish description

In order to perform our tests, we selected a total of 12 specimens of jellyfish provided by Oceanogràfic de Valencia. All the individuals belong to *C. mosaicus* and have a similar age. In this case, even though they were all the same species, they presented two types of colouration (blue or brown). We balanced the number of brown and blue specimens among the selected individuals. Their size, measured using a calliper, varies between 2 cm and 5 cm, the smallest 2,66 cm and the largest 4,87 cm, as shown in Table 1.

D. Conducted test

The methodology followed for this experiment can be separated into two stages. In all the tests, the coil is placed in the aquarium. The first one, where the EM field was not activated, also known as the blank test, serves to have information about the regular motion of jellyfish. In the second test, also known as the exposition test, the EM field was activated by powering the coil with the generator. This simulates the exposition of the proposed system. Each stage consisted of a 5 minutes lapse in which the behaviour was recorded every minute.

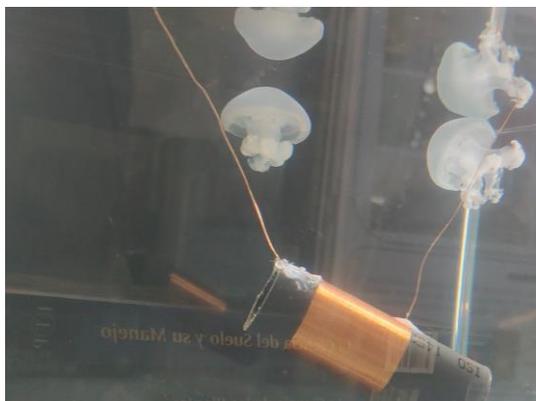


Figure 1. Assembled test with jellyfish and the coil.

TABLE I. SUMMARY OF CHARACTERISTICS OF THE SELECTED JELLYFISH

ID	Characteristics	
	Size (cm)	Colour
1	2,78	Blue
2	3,26	Brown
3	4,06	Blue
4	4,52	Brown
5	3,85	Blue
6	3,71	Brown
7	3,39	Brown
8	4,63	Brown
9	3,73	Blue
10	4,87	Blue
11	3,61	Blue
12	3,24	Brown

E. Behaviour's indicator

The pulse rate corresponds to the number of complete pulses performed by the jellyfish every 30 seconds. The pulses were counted during blank and exposition tests. The pulse rate has been selected as an indicator to characterise the jellyfish's behaviour.

Measurements were taken in slots of time of 30 seconds along the 5 minutes of trials. The first data collection started 30 seconds after the beginning of the experiment, and the last one was at 4 minutes and 30 seconds.

IV. RESULTS

In this section, we describe the obtained results from the conducted tests. First, the results of the descriptive statistics are shown. The comparison of jellyfish behaviour when they are exposed or not to the EM fields is detailed later. Finally, the evaluation of the suitability of this system to prevent jellyfish on our shores is discussed.

A. Descriptive statistics of obtained results

Next, we describe the variability of data gathered during the different experiments. First of all, five histograms are included in Figure 2, in which we can see the distribution of pulses. During the data collection in the tests, we noted that the blue jellyfish behave differently than the brown ones. Thus, besides the histograms for the EM field and no EM field, the data is also divided according to the colour of the jellyfish. In addition, the descriptive analyses of this data are summarised in Table 2.

In Figure 2 a), we can see the histogram for all collected data; it is possible to identify that the variable does not follow a normal distribution (kurtosis and skewness in Table 2). In most cases, jellyfish pulse rate are close to 50 pulses every 30 seconds. Nonetheless, we can see differences when differentiating between colours, especially when jellyfish are exposed to the EM field. The blue jellyfish have, in most cases, values close to 50 pulses every 30 seconds, see Figure 2 b). When they are exposed, the histogram shows a decrease in the number of pulses, see Figure 2 c). Concerning the brown individuals (Figure 2 d)), even when they are not exposed to the EM field, they have a lower pulse rate, with the maximum of the histogram in 40. Finally, when brown individuals are exposed to the EM field, they decrease the

number of pulses even more. It is possible to find pulse rate values below 20, see Figure 2 e).

When data is analysed independently according to the jellyfish colour and EM field exposition, in some cases, the pulse rate follows a normal distribution. This happens with data on brown jellyfish.

The main conclusion of the descriptive analyses is that if all data is going to be analysed, non-parametric methods must be used. In addition, the descriptive analyses have already shown that there are differences between individuals when they are exposed to the EM field. These differences are more evident for brown jellyfish than for blue ones.

B. Comparison of jellyfish behaviour when they are exposed to EM field

Now, the comparison of results among tested individuals when they are exposed to EM fields is presented. Figure 3 and Figure 4 depict the registered motion of jellyfish under the blank and exposition test. The X-axis includes the

number of individuals and the type of exposition (Yes or Not). It is possible to see that the pulsation rate along the test is similar for the blue individuals, with very few differences between individuals and in exposition. For the brown individuals, higher variability in data is seen. It is possible to see in Figure 5 that when individuals are under EM, the pulses decrease considerably.

In order to evaluate if differences are statistically significant, we used a Kruskal-Wallis non-parametric test since parameter did not follow a normal distribution. Table 2 summarises the results of the different tests. First, we compared the data of all individuals using the presence of the EM field as a factor. The results indicate that differences were not statistically significant (p-value of 0.098). The case in which differences in motion due to the presence of EM field were statistically significant was for the brown individuals, with a p-value of 0.02.

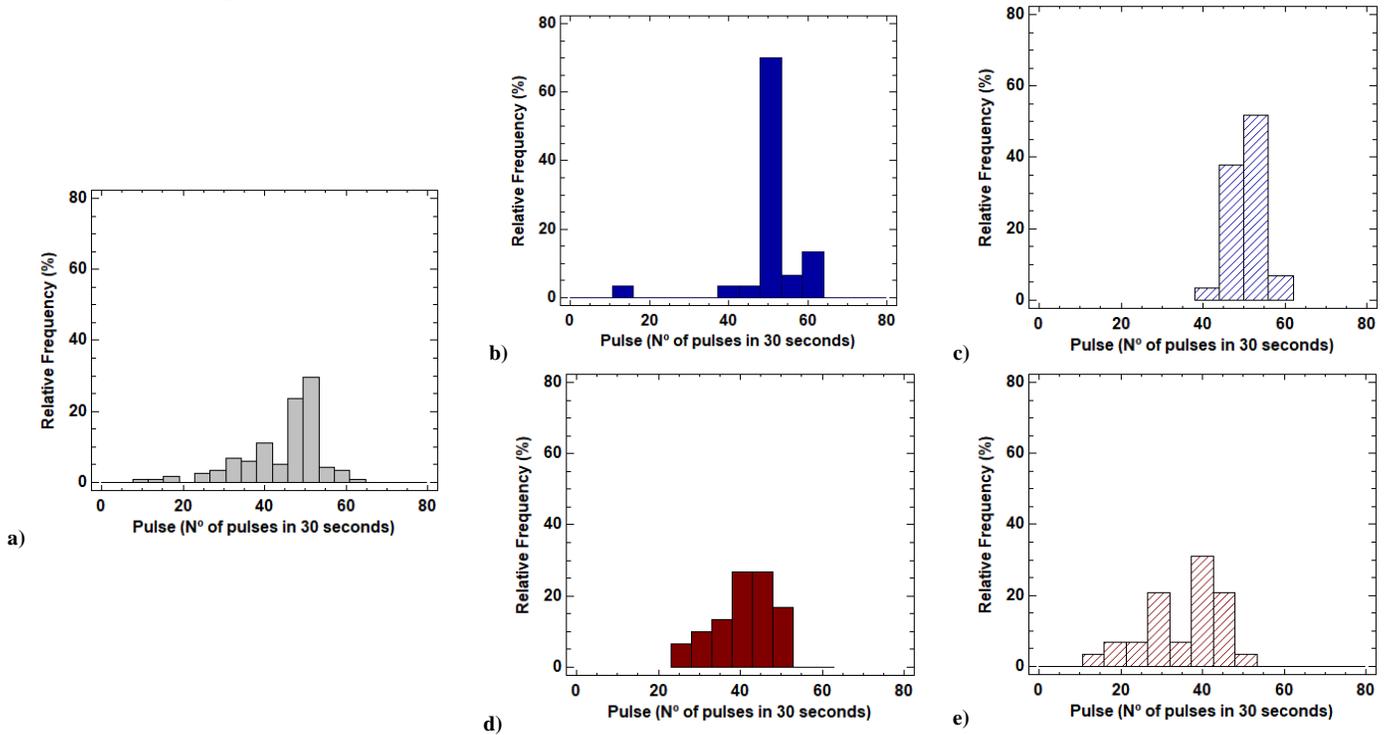


Figure 2. Histograms of pulses recorded in the different experiments including a) all data, b) blue jellyfish with no EM field, c) blue jellyfish with EM field, d) brown jellyfish with no EM field, and e) brown jellyfish with EM field.

TABLE II. SUMMARY OF CHARACTERISTICS OF THE SELECTED JELLYFISH

ID	Used data				
	All	Blue with no EM	Blue with EM	Brown with no EM	Brown with EM
N	118	30	29	30	29
Mean (Pulses/30s)	44.66	50.96	50.55	41.27	35.75
σ	9.89	8.63	3.77	7.12	9.57
Kurtosis	-5.14	-8.10	-1.47	-1.35	-1.47
Skewness	2.78	18.58	2.38	-0.45	-0.26

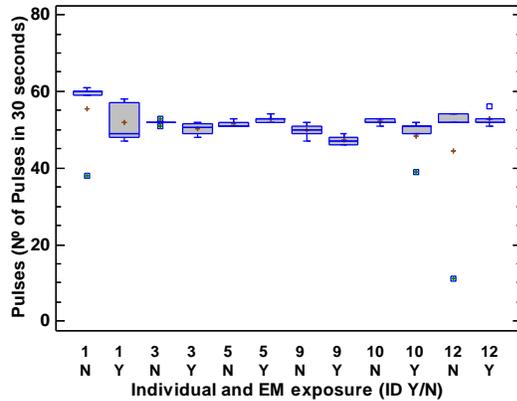


Figure 3. Boxplot for blue individuals during exposition tests (Y) and in blank tests (N).

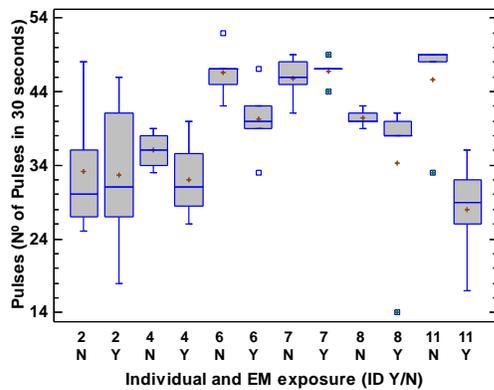


Figure 4. Boxplot for brown individuals during exposition tests (Y) and in blank tests (N).

TABLE III. SUMMAY OF KRUSKAL-WALLIS TESTS

ID	Factor	Characteristics	
		Statistic	p-value
Brown individuals	EM field	5.32949	0.0209648
Blue individuals	EM field	2.64529	0.103853
All individuals	EM field	2.72585	0.0987317

C. Impact and limitations

Our results indicate that using an EM field only for 5 minutes can alter the mobility of brown jellyfish. During the test, the jellyfishes that stopped pulsing were in the bottom of the aquarium. Regarding the blue individuals, even though the pulse rate was not altered, they experienced moments when they stopped pulsing and went to the bottom, as can be seen in Figure 5. After each period in which the jellyfish stopped pulsing when the jellyfish start pulsing the pulse was faster than before. This explains why in some individuals, even if they remain for some seconds without pulsing, the pulse rate does not vary. Nevertheless, in those situations, the velocity of the jellyfish was almost null. Although an exhaustive study of the jellyfish's behaviour has not been carried out after its exposure to the EM field, it is possible to see a partial recovery in their ability to pulsate.



Figure 5. Picture of jellyfish in the bottom of the aquarium due to the EM.

Before applying this solution to real environments, some adjustments must be made. The system can be extended by adding a method to detect the jellyfish for activating the EM field. It is not expected any problem between affected and non-affected jellyfish since, as far as we know, no communication has been described among jellyfish. Moreover, during the performed test, it was not detected that the jellyfish avoided contact with the coil that generated the EM field.

The most challenging issues of this system are the powering and the structural aspects. In addition, it will be necessary to study the potential effect of generated EM fields on other organisms. The full system is expected to be developed and tested in the next year thanks to the collaboration of partners of the SALVADOR project.

V. CONCLUSIONS AND FUTURE WORK

The arrival of jellyfish blooms to the coast significantly affects areas in which the economy relies on tourism. Existing any-jellyfish barriers, mainly based on nets or air bubbles, are not fulfilling the expectations. Some public agencies are demanding adequate jellyfish management strategies [19].

In this paper, we presented a new any-jellyfish system based on the generation of EM fields that modify jellyfish's motion. The impact of EM on mobility was tested with 12 individuals of *Catostylus mosaicus*. A copper coil and an alternating current generator generated the EM field. The results show that EM fields alter the mobility of half of the tested individuals according to the registered pulse rate. Initial results pointed out that using the EM field has clear potential as an alternative option for an anti-jellyfish barrier.

The impact of these EM fields in other species of jellyfish as well as in other local fauna, must be tested in future work. Additional tests will include changing the signal used to power the copper coil to evaluate the effect of different generated EM fields. It would be interesting to study the behaviour in adult specimens by using the same parameters in this study in order to compare behaviours with

younger individuals. Furthermore, since it is intended to modify the used signal, it can be tested in adult jellyfish. Another possible work for the future would be to change the size and type of coil, using a larger one or changing the number of spires to generate an EM field with higher magnetic flux density.

ACKNOWLEDGMENT

This research was funded by ThinkInAzul programme and was supported by MCIN with funding from European Union NextGenerationEU (PRTR-C17.I1) and by Generalitat Valenciana (THINKINAZUL/2021/002), by “Proyectos de innovación de interés general por grupos operativos de la Asociación Europea para la Innovación en materia de productividad y sostenibilidad agrícolas (AEI-Agri)” in the framework “Programa Nacional de Desarrollo Rural 2014–2020” project GO TECNOGAR, and by the “Generalitat Valenciana” through the project INVEST/2022/467. It had also support from the Ajuntament de Dénia, Conselleria d’Agricultura, Desenvolupament Rura, Emergència Climàtica i Transició Ecològica de la Generalitat Valenciana and the University of Alicante through the Montgó-UA-Dénia Research Station agreement. We want to thank the Oceanografic staff for providing the jellyfish.

REFERENCES

- [1] B. Mghili, M. Analla, and M. Aksissou, “Estimating the economic damage caused by jellyfish to fisheries in Morocco,” *African Journal of Marine Science*, 44(3), pp. 271-277, 2022, doi:10.2989/1814232X.2022.2105949.
- [2] R. E. Welton, D. J. Williams, and D. Liew, “Injury trends from envenoming in Australia, 2000-2013,” *Internal medicine journal*, 47(2), pp. 170-176, 2017, doi:10.1111/imj.13297.
- [3] M. Bosch-Belmar et al., “Jellyfish impacts on marine aquaculture and fisheries,” *Reviews in Fisheries Science and Aquaculture*, 29(2), pp. 242-259, 2021, doi:10.1080/23308249.2020.1806201.
- [4] J.E. Cloern et al., “Human activities and climate variability drive fast-paced change across the world’s estuarine-coastal ecosystems,” *Glob Change Biol*, 22(2), pp. 513-529; doi:10.1111/gcb.13059.
- [5] C. Lucas, S. Gelcich, and Shin-ichi Uye, “Living with Jellyfish: Management and Adaptation Strategies,” *Jellyfish Blooms*, pp. 129–150, 2013. doi:10.1007/978-94-007-7015-7_6.
- [6] J.M. Vasslides, N. L. Sassano, and S. Hales, “Assessing the effects of a barrier net on jellyfish and other local fauna at estuarine bathing beaches,” *Ocean and Coastal Management*, 163(), pp. 364-371, 2018, doi:10.1016/j.ocecoaman.2018.07.012.
- [7] H. G. Hiscock, H. Mouritsen, D. E. Manolopoulos, and P. J. Hore, “Disruption of magnetic compass orientation in migratory birds by radiofrequency electromagnetic fields,” *Biophysical Journal*. 113(7), pp. 1475-1484, 2017, doi:10.1016/j.bpj.2017.07.031.
- [8] V. Mildaziene et al., “Response of perennial woody plants to seed treatment by electromagnetic field and low - temperature plasma,” *Bioelectromagnetics*, 37(8), pp. 536 - 548, 2016, doi:10.1002/bem.22003.
- [9] D. Haberlin, R. McAllen, and T. K. Doyle, “Field and flume tank experiments investigating the efficacy of a bubble curtain to keep harmful jellyfish out of finfish pens,” *Aquaculture*, 531, 735915, 2021, doi:10.1016/j.aquaculture.2020.735915
- [10] N. Killi at al., “Risk screening of the potential invasiveness of non-native jellyfishes in the Mediterranean Sea,” *Marine Pollution Bulletin*, 150, 110728, 2019, doi:10.1016/j.marpolbul.2019.110728.
- [11] T. A. Morsy, N. M. Shoukry, and M. A. Fouad, “Jellyfish Stings: Complications And Management,” *Journal of the Egyptian Society of Parasitology*, 50(2), pp. 270-280, 2020, doi:10.21608/jesp.2020.113048.
- [12] Z. L. Hutchison, A. B. Gill, P. Sigray, H. He, and J. W. King, “Anthropogenic electromagnetic fields (EMF) influence the behaviour of bottom-dwelling marine species,” *Scientific reports*, 10(1), 4219, 2020, doi:10.1038/s41598-020-60793-x.
- [13] B. Taormina, et al., “Impact of magnetic fields generated by AC/DC submarine power cables on the behavior of juvenile European lobster (*Homarus gammarus*),” *Aquatic Toxicology (Amsterdam, Netherlands)*, 220, 105401, 2020, doi:10.1016/j.aquatox.2019.105401.
- [14] D. P. Fey et al., “Are magnetic and electromagnetic fields of anthropogenic origin potential threats to early life stages of fish,” *Aquatic toxicology (Amsterdam, Netherlands)*, 209, pp. 150–158, 2019, doi:10.1016/j.aquatox.2019.01.023.
- [15] A. V. Kuznetsov, O. N. Kuleshova, A. Y. Pronozin, O. V. Krivenko, and O. S. Zavyalova, “Effects of low frequency rectangular electric pulses on Trichoplax (Placozoa),” *Marine Biological Journal*. 5(2), pp. 50-66, 2020, doi:10.21072/mbj.2020.05.2.05.
- [16] Web page of the current generator used. Available at: <https://www.mouser.com/datasheet/2/403/AFG1022-Arbitrary-Function-Generator-Datasheet-1-540840.pdf> [Last access on 16/02/2022].
- [17] S. Sendra, J. M. Jimenez, L. Garcia, and J. Lloret, “LoRa-based network for water quality monitoring in coastal areas,” *Mobile Networks and Applications*, 1-17, 2022, doi:10.1007/s11036-022-01994-8.
- [18] L. Parra, S. Viciano-Tudela, D. Carrasco, S. Sendra, and J. Lloret, “Low-Cost Microcontroller-Based Multiparametric Probe for Coastal Area Monitoring,” *Sensors*, vol. 23, no. 4, p. 1871, 2023, doi:10.3390/s23041871.
- [19] I. Shepherd. European efforts to make marine data more accessible. *Ethics in Science and Environmental Politics*, vol. 18, p. 75-81, 2018.