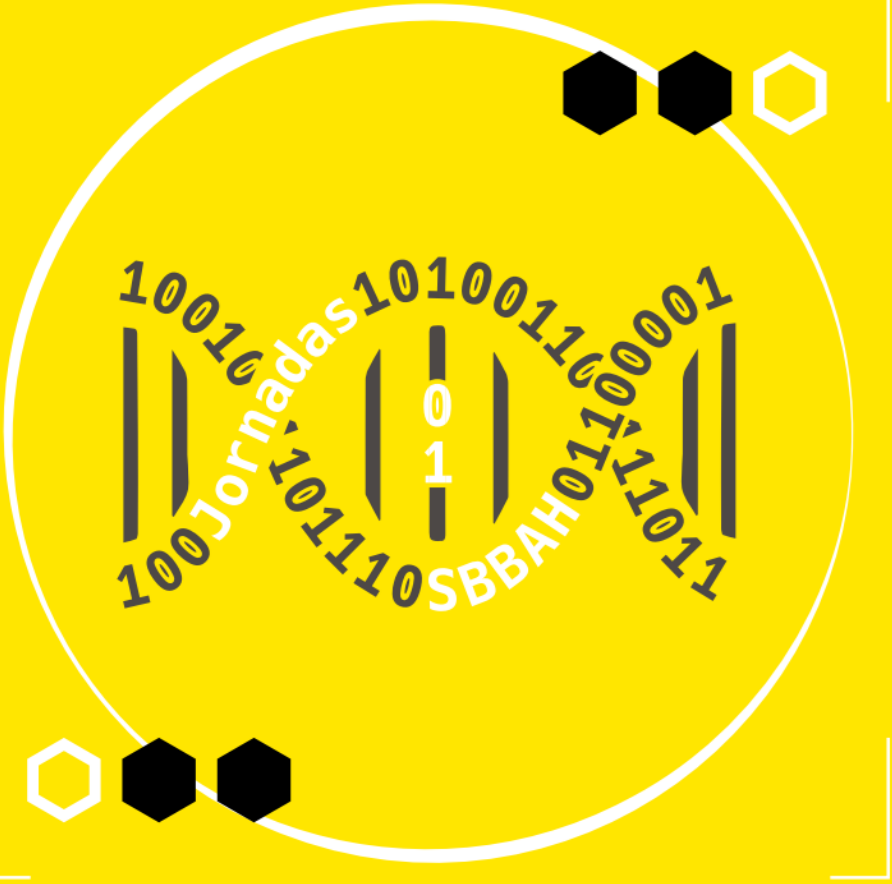
SBBAH'24



Proceedings of the 1st Symposium on Biostatistics and Bioinformatics Applied to Health

May 3, 2024, Porto, Portugal

Editors

Brígida Mónica Faria

João Paulo Martins

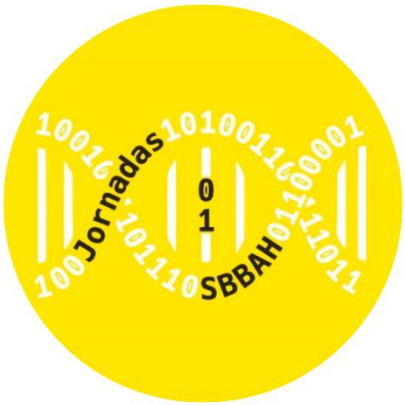
Sandra Maria Alves

Rui Pimenta

HOSPITAL DAS FORÇAS ARMADAS

P.PORTO

ESCOLA SUPERIOR DE SAÚDE



First Symposium on Biostatistics and Bioinformatics Applied to Health
//Primeiras Jornadas em Bioestatística e Bioinformática Aplicadas à Saúde

Editors

Brígida Mónica Faria

João Paulo Martins

Sandra Maria Alves

Rui Pimenta

Title: Proceedings of the 1st Symposium on Biostatistics and Bioinformatics Applied to Health

Event date: 3 May 2024

Web: https://sites.google.com/view/sbbah2024

Publication data: 14 April 2025

ISBN: 978-989-9045-35-4

DOI: https://doi.org/10.26537/edp.porto.144

Publisher: ESS | P. PORTO Edições

Preface

The proceedings of the First Symposium on Biostatistics and Bioinformatics Applied to Health (1st SBBAH) highlight the convergence of innovation and health sciences, presenting research and developments in the field. On May 3, we welcomed participants for an engaging day of discussions, beginning with an opening session that set the stage for the day's presentations.

The Symposium was organized with a diverse set of oral communications, including students and Alumni from the Master on Biostatistics and Bioinformatics Applied to Health (MBBAS) program, who presented their findings on topics ranging from machine learning applications in healthcare to the functional activity of astrocytes. An invited talk on Natural Language Processing and the role of Artificial Intelligence in Healthcare emphasized the importance of these technologies in shaping the future of health research.

We extend our gratitude to all speakers, participants, companies and organizing committees for their commitment and contributions, making this Symposium a significant way for knowledge exchange and collaboration in health, biostatistics, bioinformatics, data science and artificial intelligence.

We also would like to express our heartfelt gratitude to the MBBAS first year students who invented the symposium logo, showcasing their creativity and dedication. Additionally, we extend our thankfulness to Hospital das Forças Armadas – Polo Porto for generously hosting us in their facilities and for their sponsorship, which greatly contributed to the success of this event. Their support highlights the collaboration between academia and healthcare institutions, promoting innovation in health and science.

Conference Committees

The success of this conference was made possible by the dedication and expertise of its organizing committees, including:

General Chair

Brígida Mónica Faria

Organizing and Scientific Committees

Ana Paula Nascimento

Alexandra Alves Oliveira

Brígida Mónica Faria

Janete Borges

João Paulo Martins

Paulo Gomes Veloso

Rui Pimenta

Sandra Maria Alves

Vítor Sá

Local Chair

Carlos Lopes

Keynote Speaker

Henrique Lopes Cardoso

Talk: Natural Language Processing and Large Language Models

Henrique Lopes Cardoso holds a PhD in Informatics Engineering and an MSc in Artificial Intelligence and Computation, both from the University of Porto. He is an Associate Professor at the Faculty of Engineering of the University of Porto. He is a senior researcher and member of the Directive Board of the Artificial Intelligence and Computer Science Laboratory (LIACC). Henrique has played a significant role in the Portuguese Association for Artificial Intelligence, being involved in its Directive Board for ten years. With over 25 years of teaching experience in higher education, he has lectured on various courses within the scientific domain of intelligent systems, including Artificial Intelligence, Machine Learning, Multi-Agent Systems, and Natural Language Processing. His primary research focus is on Natural Language Processing, and he has actively participated in numerous national and international research projects in this field.

In his keynote "Natural Language Processing and Large Language Models", Professor Henrique Lopes Cardoso shared insights from his extensive research and experience, discussing the latest advancements and challenges in the field. His talk provided valuable perspectives on integrating and applying large language models in natural language processing.

Table of Contents

1. Introduction and Program Schedule

The 1st SBBAH symposium successfully brought together 78 participants, including students, Alumni, and professionals, to discuss innovative research in health sciences, data analysis, natural language processing and artificial intelligence. A total of **19 abstracts were submitted**, from which **7 were selected for oral communications** and **6 for poster presentations**. The event included a collection of oral communications from current MBBAS students and alumni, alongside invited talks by experts in the field. Participants shared their latest findings, exchanged ideas, and explored practical applications.

Below is the program schedule for the 1st SBBAH symposium:

09:30–09:45 – Opening Session
09:45–11:00 – Oral Communications (MBBAS Students) – session chairs: Rui Pimenta and João Paulo Martins
11:00–11:30 – Coffee Break
11:30–12:00 – Oral Communications (Alumni) – session chair: Sandra Alves
12:00–14:00 – Lunch Break
14:00–15:00 – Invited Talk: "Natural Language Processing and Large Language Models" – session chair: Brígida Mónica Faria
15:00–15:45 – Oral Communications (Companies/Research Laboratories) – session chairs: Paulo Veloso and Ana Paula Nascimento
15:45–16:15 – Coffee Break and Poster Session
16:15–16:50 – Oral Communications (Companies/Research Laboratories) – session chairs: Janete Borges and Vítor Sá
16:50–17:00 – Closing Session

2. Abstracts and Oral Communications: MBBAS Students

The set of oral communications presented at this event showcased a diverse range of research, highlighting the application of computational and biostatistical techniques across various fields within the health and biological sciences. Topics included advanced statistical and computational methods, such as time series forecasting for health indicators using clustering analysis and the critical role of randomization in clinical trials. Contributions to the life sciences were demonstrated through the development of a new Biopython library for molecular biology, the application of machine learning to life cycle assessment of fruit crops, the optimization of surgical scheduling using predictive models, and AI-driven swallowing assessments in clinical contexts. Additionally, the intersection of health and technology was exemplified by research exploring coping profiles of dementia caregivers using e-Health platform data and by studies predicting treatment response in exudative age-related macular degeneration through OCT biomarkers. These contributions reflect the dynamic integration of computational, statistical and practical approaches in addressing real-world challenges.

2.1. Comparing Time Series Forecasting Models for Health Indicators: A Clustering Analysis Approach (OC)

Comparing Time Series Forecasting Models for Health Indicators: A Clustering Analysis Approach

Cláudia Vinhal[1*], Alexandra Oliveira[1,2], Brígida Faria[1,2], Ana Paula Nascimento[1,3], Rui Pimenta[1,4]

1. ESS, Polytechnic of Porto, Portugal; 2. Artificial Intelligence and Computer Science Laboratory (LIACC member of LASI), Porto, Portugal; 3. Center for Translational Health and Medical Biotechnology Research (TBIO), Porto, Portugal; 4. Centre for Health Studies and Research of the University of Coimbra/Centre for Innovative Biomedicine and Biotechnology (CEISUC/CIBB), Coimbra, Portugal.

* Corresponding author email: 10220888@ess.ipp.pt

Introduction: Time series are the sequence of observations ordered by equal time intervals, crucial for understanding causality, trends, and forecasts. Its analysis can be applied to several areas, such as engineering, finance, and health (1,2). One problem with the time series study is clustering, mainly understanding when two parametric time series are considered similar (3). The sum of mortality and morbidity, referred to as "Burden of Disease", is measured by a metric called "Disability Adjusted Life Years" (DALYs) (4). These indicators are direct measures of health care needs, reflecting the global burden of disease in the population, and are crucial for public health study and surveillance (5). DALYs can be represented by Autoregressive Integrated Moving Averages (ARIMA) models, and in this context understanding clusters is crucial. **Objectives:** The primary goal is to compare different distance measures between ARIMA processes when used in clustering techniques. **Methods:** The study begins by exploring the temporal characteristics of DALYs, highlighting underlying patterns and trends. Then, ARIMA models are applied to represent and describe the time series. It's on this representation of the time series that the Piccolo, the Maharaj, and the LPC distance measures are applied to use clustering techniques and identify clusters. Additionally, 8 distinct cluster validation metrics are used. **Results:** Specific to 48 European countries, the results show that the choice of distance measure can greatly influence clustering outcomes and the number of clusters formed. While certain methods revealed geographic patterns, other factors, such as cultural or economic similarities, also influence cluster formation. These insights contribute to advancing the field of public health surveillance and intervention, ultimately aiming to alleviate the global burden of disease. **Conclusions:** This study offers insights into applying ARIMA processes in clustering techniques for analysing temporal health data. By comparing different distance measures, this research improves our understanding of underlying patterns and trends in health indicators over time.

Keywords: Distance Measures, Clustering, DALYs, ARIMA Models

References:

1. Palma W. Time Series Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2016.

2. Nielsen A. Practical Time Series Analysis. 1st Edition. Sebastopol, CA: O'Reilly Media Inc.; 2019.
3. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering – A decade review. Inf Syst. 2015 May 30;53:16–38.
4. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019). Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2020.
5. Mathers CD. History of global burden of disease assessment at the World Health Organization. Archives of Public Health. 2020 Aug 24;78(1).

2.2. New Biopython Library to Support Molecular Biology (OC)

New Biopython Library to Support Molecular Biology

Patrícia Nogueira[1], Vítor Sá[1]

1. ESS, Polytechnic of Porto, Porto, Portugal

10190914@ess.ipp.pt; vitor.sa@ess.ipp.pt

Introduction: Biopython is a library that facilitates the development of applications for Bioinformatics, using the Python programming language. Maintained by an international association of programmers - the Open Bioinformatics Foundation, it provides tools for analysing biological sequences and accessing online databases like NCBI. It features modules for sequence alignment, protein structures, population genetics, and more. **Objectives:** Since the library is open-source and has the collaboration of several developers, the project's aim is to create a new library that will recur to existing API services such as Expasy, Blast, Uniprot and DrugBank. This will enable a single-call module to reference multiple services, obtain results, and generate a final report for a searched sequence. **Methods:** The first step to start the project was a study of existing Biopython libraries to assess their alignment with the proposed objectives. Access to the DrugBank service required a formal request. It was necessary to justify the request, explaining how and why the data would be used within the scope of the project. After the preceding steps, the entire architecture and design of the solution were conceptualized. Subsequently, we started the development. **Results:** As a result of this project, there will be an endpoint that, when invoked by any software or platform, should return all information found for a genetic sequence in JSON format. Additionally, a small user interface was developed to display the search results as an alternative to using only the API. **Conclusions:** Because BioPython is a free package and science is the motivation a new library will be available several services as Expasy, Blast, Uniprot and DrugBank. Free for all schools, labs, researchers and all developers who want to use it.

Keywords: ExPASy; Blast; UniProt; DrugBank

References:

1. API Reference | DrugBank Help Center [Internet]. [cited 2023 Oct 26]. Available from: https://docs.drugbank.com/v1/
2. Biopython · Biopython [Internet]. [cited 2023 Oct 26]. Available from: https://biopython.org/
3. BLAST: Basic Local Alignment Search Tool [Internet]. [cited 2023 Oct 26]. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi
4. DrugBank | Clinical Drug Data API [Internet]. [cited 2023 Oct 26]. Available from: https://www.drugbank.com/clinical
5. Expasy - Translate tool [Internet]. [cited 2023 Oct 26]. Available from: https://web.expasy.org/translate/ 7.ushinauwu. OBF » Projects » Projects [Internet]. [cited 2023 Oct 26]. Available from: https://www.open- bio.org/projects/
6. OMIM API Help - OMIM [Internet]. [cited 2023 Oct 26]. Available from: https://www.omim.org/help/api
7. OpenAI Platform [Internet]. [cited 2023 Oct 26]. Available from: https://platform.openai.com

8. UniProt [Internet]. [cited 2023 Oct 26]. Available from: https://www.uniprot.org/

2.3. Life Cycle Assessment using Machine Learning (OC)

Life Cycle Assessment using Machine Learning

Sofia Gomes[1*], Brígida Mónica Faria[1,2], Alexandra Alves Oliveira[1,2], Edgar Pinto[1,4], Matilde Rodrigues[1,3], Manuela Vieira[1,4]

1. ESS, Polytechnic of Porto, Portugal; 2. Artificial Intelligence and Computer Science Laboratory (LIACC member of LASI), Porto, Portugal; 3. Center for Translational Health and Medical Biotechnology Research (TBIO), Porto, Portugal; 4. REQUIMTE/LAQV, ESS, Polytechnic of Porto, Porto, Portugal.

* Corresponding author email: 10221255@ess.ipp.pt

Introduction: Life Cycle Assessment (LCA) is a scientific tool that allows calculating the impact of a product or service on the environment, considering the different phases from planting to transportation, commercialization, consumption, and disposal. (1) LCA requires comprehensive data collection of the inputs and outputs such as raw materials, energy, water, used chemicals and pollutants emissions at each stage of the life cycle. Data is usually obtained from different sources like producers or farmers (primary data), literature reviews, government reports and scientific publications (secondary data) or from associations, non-governmental organizations (NGOs) and international organizations. (2) Data processing and analysis are conducted with the aim of uncovering the resultant environmental impacts. This dissertation, integrated into the project REtail using Technology based on Artificial InteLLigence (RETAILL) (3) aims to apply Machine Learning (ML) techniques to develop surrogate LCA models that can be used to estimate LCA results for new products or services. Both public data and data from the Terras de Felgueiras Cooperative (4) will be used to develop the intended model. By preprocessing and modelling this data, the study aims to provide valuable insights for enhancing sustainability in the production of fresh fruits and vegetables. These insights can guide decision-making and drive continuous improvement in the supply chain. **Objectives:** The objective of this study is to develop a ML model that estimates LCA results for new products or services and that translates environmental indicators into measurable impacts on both the environment and human health, specifically focusing on the production process. Another objective is to establish clusters that represent similar environmental performance of producers or products. **Methods:** The first step was to review the existing literature on the subject. To accurately determine emissions from agricultural activities, validated equations from the Agri-footprint 6 methodology (5) were employed. Preliminary analyses and descriptive statistics of variables such as fertilizers, pesticides and fungicides applied on agriculture from public data assessments were conducted using tools like SPSS and RapidMiner. This latter was used to carry out the construction of decision trees and clusters. To evaluate clustering models, certain indices were considered namely the Davies-Bouldin Index and the Calinski-Harabasz Index. Meanwhile, for assessing decision trees, measures such as accuracy rate, F- measure, and confusion matrix are well-known evaluation criteria. Subsequently, the ML model was developed using Python programming language and libraries such as Scikit- learn, Pandas, and SciPy. **Results:** The analysis of public data reveals results from the cultivation of kiwi, watermelon, citrus, tea, and

hazelnut across a total of 865 orchards (6). The results include the development of a tailored ML model for LCA phase that allows. the identification and translation of key environmental indicators into environmental and human health impacts. Furthermore, the clustering results of products and producers enables the observation of patterns in the environmental impact of the production process. **Conclusions:** Overall, this study contributes to the field of sustainability by providing a framework for integrating ML techniques with life cycle assessment, ultimately leading to more efficient and effective practices in agricultural production. The utilization of validated equations from Agri-footprint 6 enhances the reliability of emissions determination from agriculture, contributing to more accurate assessments of environmental impacts. Ultimately, the goal of an LCA is to support informed decision- making and promote more sustainable practices across industries.

Keywords: Life Cycle Assessment, Sustainability, Machine Learning, Data analysis

References:

1. The data you need to make a Life Cycle Assessment (LCA) [Internet]. Ecochain. [cited 2024 Apr 18]. Available from: https://ecochain.com/blog/input-data-for-lca/
2. Pergola M, Persiani A, D'Ammaro D, Pastore V, D'Adamo C, Palese AM, et al. Environmental and Energy Analysis of Two Orchard Systems: A Case Study in Mediterranean Environment. Agronomy. 2022 Oct;12(10):2556.
3. RETAILL's Consortium. REtail using Technology based on Artificial InteLLigence. 2022.
4. Cooperativa Terras de Felgueiras. Apresentação [Internet]. 2020 [cited 2024 Jan 17]. Available from: https://coopfelgueiras.pt/instituicao/apresentacao/
5. Sustainability B. Agri-footprint 6 Methodology Report.
6. Mostashari-Rad F, Ghasemi-Mobtaker H, Taki M, Ghahderijani M, Kaab A, Chau K wing, et al. Exergoenvironmental damages assessment of horticultural crops using ReCiPe2016 and cumulative exergy demand frameworks. J Clean Prod. 2021 Jan 1;278.

2.4. Optimization of Surgical Scheduling: Predicting Surgery Time Duration using Machine Learning (OC)

Optimization of Surgical Scheduling: Predicting Surgery Time Duration using Machine Learning

Soraia Malheiro [1*] Brígida Faria [1,2] Celeste Dias [3]

1. Escola Superior de Saúde do Instituto Politécnico do Porto, Porto, Portugal; 2. Laboratório de Inteligência Artificial e Ciência de Computadores (membro do LASI); 3. Unidade Local de Saúde São João, Porto, Portugal

* Corresponding author email: 10220101@ess.ipp.pt

Introduction: The operating room (OR) is a highly specialized hospital department that requires a large amount of resources which has a high impact on hospital funding (1). The OR is an essential area for the hospital operation and its management must guarantee the best efficiency and the highest quality of patient care. Despite some initiatives already implemented to meet the demand for surgical treatment, as described in the European Commission's 2021 report (2), waiting times for surgery in Portugal have increased in the last decade. Surgical scheduling is fundamental in the OR management (3). One of the challenges related to surgical scheduling is the prediction of surgery duration, which is essential for the allocating OR occupancy times.

Currently, this prediction is essentially based on the surgeon's experience in a particular surgical procedure and may not take other variables into account (1).
Objectives: The aim of this study is to predict more accurately the duration of the surgeries in the specialties of General Surgery, Orthopedics and Urology by developing a model based in machine learning techniques with data from clinical records of surgical cases.
Methods: The sample of this study includes data from surgical cases performed in a hospital center. The following cases were excluded: Surgeries with patients under the age of 18; without a defined preoperative diagnosis; unspecified surgical specialties; no record of the start and/or end time of surgery and surgeries that took place on an outpatient basis. Multiple Linear Regression (MLR) and Random Forest (RF) techniques were applied to develop the model.
Conclusions: Accuracy in predicting the duration of surgeries can optimize the OR occupancy and at the same time decrease the waiting time experienced by the patients.

Keywords: Operating Room efficiency; Operating Room scheduling; Surgery scheduling; Machine Learning.

References:

1. Barket MA, Saxena RC, Solomon S, Fong CT, Behara LD, Venigandla R, Velagapudi K, Lang JD, Nair BG. Improving Operating Room Efficiency: Machine Learning Approach to Predict Case-Time Duration. Journal of the American College of Surgeons.2019;229(4):346-354e3. https://doi.org/10.1016/j.jamcollsurg.2019.05.029
2. OCDE. Portugal: Perfil de Saúde do País 2021, Estado da Saúde na UE, OCDE, Paris/Observatório Europeu dos Sistemas e Políticas de Saúde. 2021.
3. Lopes J, Vieira G, Veloso R, Ferreira S, Salazar M, Santos M. Optimization of Surgery Scheduling Problems Based on Prescriptive Analytics: Proceedings of the 12th International Conference on Data Science, Technology and Applications. 2023:474–479. https://doi.org/10.5220/0012131700003541

2.5. Exploring Coping Profiles in Informal Caregivers of People with Dementia through Data from an e-Health Platform (OC)

Exploring coping profiles in informal caregivers of people with dementia through an online eHealth platform

Tânia Gouveia[1*], João Viana[2], Brígida Faria[1,3], Soraia Teles[2]

1. ESS, Polytechnic of Porto, Portugal; 2. Center for Health Technology and Services Research, Porto, Portugal; 3. Artificial Intelligence and Computer Science Laboratory (LIACC member of LASI), Porto, Portugal.

*Corresponding author email: 10221256@ess.ipp.pt

Background: Informal caregivers (IC) of people with dementia (PwD) handle multiple stressors with an acknowledged impact on their physical and mental health. health (1,2). Coping strategies may be viewed as the techniques employed by ICs to minimize distress associated with the caregiving experience. iSupport emerges as an online self-guided program developed by the World Health Organization (3) to provide support and training for ICs of PwD to minimize the

negative impact of caring for a PwD. This program has been translated and culturally adapted for the Portuguese population and tested for its usability and feasibility (4,5).iSupport-Portugal[1] is, being explored for its potential as an intervention-research tool, offering support to ICs of PwD but also serving as a remote measurement tool to collect data on dementia care dyads. **Research goal**: To describe coping styles and facets among IC of PwD and to identify profiles of IC in relation to coping styles based on underlying data pattens from demographic, clinical and psychosocial variables pertaining to both the IC and PwD. **Method**: This study follows an observational cross-sectional design. Participants are IC registered on iSupport- Portugal between February 2023 and February 2024, who meet eligibility criteria. Data on the IC and PwD are collected upon registration via user's accounts. **Data analysis:** Unsupervised learning methods, such as cluster analysis, enable the identification of latent patterns or underlying structures in dimensional data. Their ability to discover similarities and differences in information makes them an ideal solution for exploring IC profiles in relation to coping strategies. **Expected results and Implications:** Coping is a modifiable dimension with known implications to well-being outcomes for both the IC and the PwD. Identifying IC profiles in relation to coping strategies is fundamental to design more targeted interventions addressing the specific needs of the IC-PwD dyads.

Keywords: Dementia, Informal care, Coping, Platform data, Cluster

References:

1. Barbosa F, Matos AD. Informal support in Portugal by individuals aged 50+. Eur J Ageing. 2014 Oct 9;11(4):293–300.
2. Iavarone A, Ziello AR, Pastore F, Fasanaro AM, Poderico C. Caregiver burden and coping strategies in caregivers of patients with Alzheimer's disease. Neuropsychiatric Disease and Treatment. 2014 Jul 29;10:1407–13.
3. OMS. Global action plan on the public health response to dementia 2017 - 2025 [Internet]. 2017 [cited 2023 Dec 19]. Available from: https://www.who.int/publications- detail-redirect/global-action-plan-on-the-public-health-response-to-dementia-2017 2025
4. Teles S, Napolskij MS, Paúl C, Ferreira A, Seeher K. Training and support for caregivers of people with dementia: The process of culturally adapting the World Health Organization iSupport programme to Portugal. Dementia. 2021 Feb;20(2):672–97.
5. Teles S, Ferreira A, Paúl C. Feasibility of an online training and support program for dementia carers: results from a mixed-methods pilot randomized controlled trial. BMC Geriatr. 2022 Dec;22(1):173.

Swallowing assessment in a clinical context: artificial intelligence applications

Teresa Torres[1,*], Brígida Faria[1,2], Pedro Araújo[1], Alexandra Oliveira[1,2], Elisa Alves[1]

1. ESS, Polytechnic of Porto, Portugal; 2. Artificial Intelligence and Computer Science Laboratory (LIACC member of LASI), Porto, Portugal.

* Corresponding author email:10160505@ess.ipp.pt

Introduction: Eating and swallowing are intricate actions comprehending both voluntary and reflexive movements, engaging over thirty nerves and muscles (1). Dysphagia occurs when the normal swallowing process is compromised, increasing the risk of the swallowed material entering the larynx, and causing complications for the patient (2). Cervical auscultation (CA) is a clinical method used to evaluate the pharyngeal phase of swallowing by listening to the sounds of swallowing-related respiration. However, the reliability of CA is susceptible to the subjectivity and experience of the speech therapist (3). The application of artificial intelligence (AI) tools in healthcare has the potential to support healthcare workers with a variety of tasks. It can be used for disease prediction and diagnostics treatment, outcome prediction and prognosis evaluation (4). **Objectives:** Contribute to the development of an AI tool to aid speech therapists in their daily evaluation of deglutition in patients, applying machine learning algorithms to process and analyse the sound recorded during cervical auscultation. **Methods:** Using an electronic stethoscope, audio samples were recorded from individuals with and without pathology while they swallowed different liquid quantities (5 mL and 10 mL) and consistency (moderately thick and solid). The audio is divided into three sections (5) and pre-processed to remove unnecessary noise. Data is classified using machine learning algorithms and the models will be evaluated according to precision, accuracy, sensibility, confusion matrix and ROC curve. **Results:** Previous work in this field reported by Santoso et al. (6) using supervised machine learning algorithms obtained promising results in swallowing detection. In this study, our dataset is composed of 87 samples of patients, and similar results are to be expected. **Conclusions:** AI can offer an objective and quantitative analysis of swallowing sounds, potentially providing more accurate and consistent results compared to subjective assessments. Additionally, it can identify complex patterns in swallowing sounds that may be challenging for less experienced speech therapists to recognize.

Keywords: Deglutition, Dysphagia, Artificial Intelligence, Machine Learning

References:

1. Matsuo K, Palmer JB. Anatomy and Physiology of Feeding and Swallowing: Normal and Abnormal. Phys Med Rehabil Clin N Am. 2008 Nov;19(4):691–707. Dudik JM, Coyle JL, Sejdic E. Dysphagia Screening: Contributions of Cervical Auscultation Signals and Modern Signal-Processing Techniques. IEEE Trans Hum-Mach Syst. 2015 Aug 1;45(4):465–77.
2. Lagarde ML, Kamalski DM, Van Den Engel-Hoek L. The reliability and validity of cervical auscultation in the diagnosis of dysphagia: a systematic review. Clin Rehabil. 2016 Feb;30(2):199–207.
3. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak. 2021 Dec 1;21(1):1–23.
4. Honda T, Baba T, Fujimoto K, Goto T, Nagao K, Harada M, et al. Characterization of Swallowing Sound: Preliminary Investigation of Normal Subjects. PLoS ONE. 2016;11(12):168187.
5. Santoso LF, Baqai F, Gwozdz M, Lange J, Rosenberger MG, Sulzer J, et al. Applying Machine Learning Algorithms for Automatic Detection of Swallowing from Sound. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2019. p. 2584–8.

2.7. Predicting Treatment Response in Wet Age-Relatedd Macular Degeneration Through OCT Biomarkers (OC)

Predicting Treatment Response in Wet Age-Relatedd Macular Degeneration Through OCT Biomarkers

Vânia Sousa[1,2,*], Ângela Carneiro[2,3], Brígida Mónica Faria[1,4]

1. ESS - Polytechnic of Porto, Portugal; **2**. São João Health Local Unit, Porto, Portugal; **3**. Faculty of Medicine of the University of Porto, Portugal; **4**. Artificial Intelligence and Computer Science Laboratory (member of LASI), Porto, Portugal.

* Corresponding author email: vania.sousa@ulssjoao.min-saude.pt

Introduction: Wet age-related Macular Degeneration (AMD), characterized by macular neovascularization that leads to fluid leakage and retinal hemorrhage causing severe and sometimes irreversible visual damage, is one of the biggest causes of blindness in developed countries (1). Treatment with anti-vascular endothelial growth factors (anti- VEGF) has been revolutionary, but not all patients respond completely to treatment and have unmet clinical needs (2). The schedule and response to these treatments is a burden for patients/carers and hospitals (2). The use of optical coherence tomography (OCT) has become a valuable tool for diagnosing and monitoring this pathology. There are biomarkers in the retina detectable with state-of-the-art OCT that may be associated with visual recovery after treatment with anti-VEGF (3). **Objectives:** Identify the biomarkers with the most influence on the response to treatment and develop various supervised learning algorithms to predict the response to treatment in order to better adapt the treatment to each patient, reducing the burden that the tight schedule of these injections has on patients/caregivers, hospitals and health professionals. **Methods:** We collected general data, visual acuity and OCT from patients with wet AMD undergoing treatment with anti-VEGF injections, followed at the São João Local Health Unit for 3 years. A statistical analysis and study of the variables with the greatest weight in the response to treatment will be carried out. With these variables, we intend to use

various supervised learning algorithms to see if it is possible to create a model with a good accuracy rate for predicting the response to treatment. **Results:** We have collected data from 98 eyes of 81 patients, 29 female (35.8%) and 52 male (64.2%) with mean age of 76.93 ± 7.6 years, with mean initial visual acuity of 60.15 letters and 58.76% of eyes with type I membrane. With this data we will identify the biomarkers with the most influence on the response to treatment and select the algorithm with the best model evaluation metrics. **Conclusions:** By identifying biomarkers and selecting an algorithm, we can find ways to improve patient treatment. Making this study multicentric would be an improvement, but data collection always requires specialized professionals and is time-consuming.

Keywords: Wet age-related macular degeneration, anti-VEGF treatment, ocular coherence tomography biomarkers, supervised learning algorithms

References:

1. Liberski S, Wichrowska M, Kocięcki J. Aflibercept versus Faricimab in the Treatment of Neovascular Age-Related Macular Degeneration and Diabetic Macular Edema: A Review. Vol. 23, International Journal of Molecular Sciences. MDPI; 2022.
2. Mettu PS, Allingham MJ, Cousins SW. Incomplete response to Anti-VEGF therapy in neovascular AMD: Exploring disease mechanisms and therapeutic opportunities. Vol. 82, Progress in Retinal and Eye Research. Elsevier Ltd; 2021.
3. Metrangolo C, Donati S, Mazzola M, Fontanel L, Messina W, D'Alterio G, et al. OCT Biomarkers in Neovascular Age-Related Macular Degeneration: A Narrative Review. Vol. 2021, Journal of Ophthalmology. Hindawi Limited; 2021.

2.8. Randomization in Clinical Trials: a biostatistician task

Randomization in Clinical Trials: a biostatistician task

Heitor Coelho[1,3*], João Albuquerque[1,2], Sandra Maria Alves[3,4]

1. BlueClinical- Investigação e desenvolvimento em saúde, Lda, Porto, Portugal; **2.** CEAUL – Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal; **3.** ESS, Polytechnic Institute of Porto, Porto. Portugal; **4.** Centre for Health Studies and Research of the University of Coimbra/Centre for Innovative Biomedicine and Biotechnology (CEISUC/CIBB), Coimbra, Portugal

* Corresponding author email: rafaelcoelho.160@gmail.com

Introduction: Randomized clinical trials stand as the golden standard for evidence-based clinical investigation on humans, providing robust methodologies to evaluate new interventions (1). By employing stochastic processes to allocate participants to treatment and control groups, randomization minimizes bias and enhances the validity of results (1,2). This ensures unpredictability in treatment assignment, mitigating selection, response, and confounding biases, thus safeguarding the statistical integrity of the study (1,3,4). Various randomization techniques address specific trial needs (5). Simple randomization is the most straightforward, offering equal allocation probabilities but risking imbalance in small samples (6). Block randomization ensures numerical balance across groups but may fail to address covariate comparability (7). Stratified randomization improves group balance concerning key covariates but demands careful variable

selection to prevent empty strata (5,6). Adaptive randomization adjusts allocation probabilities during the trial to manage imbalances but introduces complexity and potential predictability (1,8). Selecting the appropriate randomization method requires careful consideration of study design, sample size, and confounding factors to ensure reliable results (5). Beyond method selection, effective implementation is crucial, necessitating adherence to best practices outlined in the CONSORT 2010 guidelines (1). Proper documentation of allocation sequence generation, masking procedures, and protocol monitoring ensures reproducibility and integrity (1). This session will introduce randomization in clinical trials, present different methodologies, and explore their strengths and weaknesses.

Keywords: Clinical trials, Biostatistics, Randomization

References:

1. Pamela A. Shaw, Laura Lee Johnson, Craig B. Borkowf, John I. Gallin, Frederick P. Ognibene, Laura Lee Johnson. Chapter 23 – Issues in Randomization. In: Principles and Practice of Clinical Research (Fourth Edition). Boston: Academic Press. 2018. p. 329–39.
2. ICH E6 (R2) Good clinical practice - Scientific guideline | European Medicines Agency [Internet]. [cited 2024 Jan 22]. Available from: https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice-scientific-guideline
3. Meinzen-Derr J, Smith L. Sources of Error: Selection Bias, Information Bias, and Confounding. In: Handbook for Clinical Research: Design, Statistics, and Implementation. Demos Medical Publishing. 2014. p. 171–3.4.
4. Berger VW, Bour LJ, Carter K, Chipman JJ, Everett CC, Heussen N, et al. A roadmap to using randomization in clinical trials. BMC Med Res Methodol. 2021 Aug 16;21(1):168.
5. Kang M, Ragan BG, Park JH. Issues in Outcomes Research: An Overview of Randomization Techniques for Clinical Trials. J Athl Train. 2008;43(2):215–21.
6. Lim CY, In J. Randomization in clinical studies. Korean J Anesthesiol. junho de 2019;72(3):221.
7. Tilley B, Schork A. Clinical Trials: II. Randomization and Sample Size. Henry Ford Hosp Med J. 1985;33(4):219–224.
8. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. John Wiley & Sons; 2013.

2.9. Validação do questionário: "A survey of pharmacist knowledge, attitudes, utilization and barriers toward artificial intelligence"– Tradução e retrotradução

Contributions to the cross-cultural validation of "A survey of pharmacist knowledge, attitudes, utilization and barriers toward artificial intelligence": translation and back translation

Sofia Gerardo[1], Rui Pimenta[1,2], Sandra Maria Alves [1,2]

1. ESS, Polytechnic Institute of Porto; **2.** Centre for Health Studies and Research of the University of Coimbra/Centre for Innovative Biomedicine and Biotechnology (CEISUC/CIBB), Coimbra, Portugal

Introduction: The use of Artificial Intelligence (AI) is rapidly transforming various fields, and pharmacy is no exception. AI is increasingly being used to automate, optimize, and personalize

various tasks in pharmacy practice, from drug discovery to dispensing to patients. In Community Pharmacy, in addition to these possibilities, it allows for personalized and focused patient care through the selection of more appropriate and personalized therapies, with a lower probability of prescription errors and drug interactions, as well as monitoring of therapy adherence. Despite these potential benefits, its implementation in the pharmaceutical field, as well as in other areas of healthcare, should be carefully considered, as ethical and regulatory issues may pose obstacles. Likewise, the perspective and experience of each professional, which remain highly personal, especially in patient care, should not be overlooked. Therefore, it is increasingly important to know the knowledge, attitudes, utilization, and barriers concerning AI. Firstly, knowledge, as this concept encompasses the level of awareness and understanding that individuals or organizations have regarding AI technologies. Attitudes, which refers to the perceptions, feelings, and predispositions towards AI. It includes both positive and negative sentiments, such as excitement about AI's potential benefits, concerns about ethical implications, and fears about job displacement. Finally, the barriers or obstacles that hinder the adoption and effective utilization of AI. Barriers can be technical, such as lack of expertise or inadequate infrastructure; financial, such as high costs of implementation; or cultural, such as resistance to change or lack of trust in AI systems. **Objectives:** This study aimed to translate and validate a survey instrument designed to explore the attitudes of Community Pharmacy professionals towards the implementation of Artificial Intelligence (AI) in their field. **Methods:** According to the COSMIN methodology, the initial translation of the construct from its original language, English, into European Portuguese was executed by two independent translators possessing comprehensive understanding of the questionnaire concepts. Both translators are bilingual healthcare professionals, who perform functions in a hospital environment and with European Portuguese being their native language. In the subsequent step, the two acquired versions were juxtaposed, leading to the formation of a consensus version endorsed by specialists, considering the new context where the construct will be applied, without excluding the original version. In cases where there are discrepancies between the two translations, the expert panel discusses the alternatives and decides on the most suitable option. The questionnaire translation process culminates with back-translation, wherein the consensus version obtained is rendered back into the original language, English, by a bilingual translator. The resultant back-translation should closely mirrors the original questionnaire, signifying the efficacy of the content translation process. Additionally, reliability testing methods like test-retest reliability and internal consistency checks help verify the stability and consistency of the survey results. To assess consensus among different questionnaire versions, techniques such as inter-rater reliability, the Delphi method, and agreement indices are essential. These processes ensure that the survey items are interpreted consistently across different respondents and that any subjective judgments are reliably measured. **Results:** The results of the translated version questionnaire maintain the intended constructs and adequately capture attitudes toward AI implementation among Community Pharmacy professionals. **Conclusions:** Understanding the attitudes and perceptions of pharmacy professionals towards AI implementation is crucial for informing policy decisions, designing targeted interventions, and facilitating the successful integration of AI technologies into pharmacy practice. It is intended that this questionnaire contributes to the growing body of literature on AI in healthcare and serves as a foundation for further investigations into this evolving field. Future work includes the validation of the PT-EU questionnaire.

Keywords: Artificial intelligence, Community Pharmacy, Translation and validation

References:

1. Raza MA, Aziz S, Noreen M, Saeed A, Anjum I, Ahmed M, Raza SM. Artificial Intelligence (AI) in Pharmacy: An Overview of Innovations. Innov Pharm. 2022 Dec 12;13(2):10.24926/iip.v13i2.4839. doi: 10.24926/iip.v13i2.4839. PMID: 36654703; PMCID: PMC9836757, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9836757/.
2. Anan S. Jarab, Walid Al-Qerem, Karem H Alzoubi, Haneen Obeidat. Artificial intelligence in pharmacy practice: Attitude and willingness of the community pharmacists and the barriers for its implementation, Saudi Pharmaceutical JournalVolume 31, Issue 8, August 2023, 101700, https://www.sciencedirect.com/science/article/pii/S1319016423001950
3. Osama Khan, Mohd Parvez, The future of pharmacy: How AI is revolutionizing the industry, Intelligent PharmacyVolume1, Issue1, June 2023, Pages 32-40, https://www.sciencedirect.com/science/article/pii/S2949866X23000084
4. Marie-Fabienne Fortin, PhD, "O processo de investigação: da concepção à realização", 1991, Lusociência Edições técnicas e científicas, Lda. https://www.academia.edu/42384751/O_processo_de_investiga%C3%A7%C3%A3o_FORTIN
5. Simone Castagno, Mohamed Khalifa, "Perceptions of Artificial Intelligence Among Healthcare Staff: A Qualitative Survey Study", Front. Artif. Intell., 21 October 2020, Sec. Medicine and Public Health. https://www.frontiersin.org/articles/10.3389/frai.2020.578983/full
6. Ali Z, Bhaskar SB. Basic statistical tools in research and data analysis. Indian J Anaesth 2016; 60:662-9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5037948/
7. Tsang S, Royse CF, Terkawi AS. "Guidelines for developing, translating, an validating a questionnaire in perioperative and pain medicine". Saudi J Anaesth 2017;11: S80-9. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5463570/
8. Saraiva D., Almeida A., "Tradução e adaptação cultural do Safety Attitudes Questionnaire – Short Form 2006 para Portugal", Port J Public Health (2018) 35 (3): 145–154, https://karger.com/pjp/article/35/3/145/275025/Traducao-e- adaptacao-cultural-do-Safety-Attitudes
8. CB Terwee, PhD, VU University Medical Center, "COSMIN methodology for assessing the content validity of PROMs – Version 1", https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user- manual-v1.pdf.

2.10. Modelling therapeutic response in asthmatic adults: a previous exploratory analysis

Modelling therapeutic response in asthmatic adults: a previous exploratory analysis

Cristina Alves[1,2*], Brígida Mónica Faria[1,3], Sandra Maria Alves[1,4], Jorge Ferreira[2]

1. ESS – Polytechnic of Porto, Portugal. **2**. Local Health Unit Between Douro and Vouga, Santa Maria da Feira, Portugal; **3**. LIACC – Artificial Intelligence and Computer Science Lab (member of LASI), Porto, Portugal; **4**. Centre for Health Studies and Research of the University of Coimbra/Centre for Innovative Biomedicine and Biotechnology (CEISUC/CIBB), Coimbra, Portugal

* Corresponding author email: 10220098@ess.ipp.pt

Introduction: Asthma is a respiratory disease characterized by chronic inflammation of the airways. Effective asthma management is essentially based on choosing the appropriate treatment for each individual (1). Data science and machine learning models offer valuable insights and enhance the outcomes achieved in asthma management (2). **Objectives:** The main objective is to develop predictive models for therapy response in patients with asthma, and secondarily to identify clinical, functional and biological characteristics that influence this response. **Methods:** Data from fifty adults with asthma were analyzed, collecting information on anthropometric, clinical, functional, biological, therapeutic, occupational, and allergen exposure

factors. The study followed the "Knowledge Discovery in Databases, KDD" methodology. **Results:** The sample consisted of 50 asthmatic adult participants, aged between 21 and 81 years old mean age=54.02 (s=14.5), from which 20 (40%) were male and 30 (60%) were female. The analysis of the characteristic symptoms of asthma (dyspnea, cough, wheezing and chest tightness), reveals a statistically significant improvement ($p<0.001$) of all these symptoms after the treatment. The asthma control test, the life quality questionnaire and the asthma and allergic rhinitis control test evaluated before and after treatments, demonstrate a statistically significant difference ($p=0.023$, $p<0.001$ and $p<0.001$, respectively). On respiratory function, only FVC reveals a significant difference ($p=0.409$), after treatment. However, the average did not reach the minimal important difference (MID) of 200ml. The average number of exacerbations and SU recurrences difference was also significant in both cases ($p<0.01$), reaching MID (>50%). **Conclusions:** The majority of the individuals in this group had a positive, clinically important response to treatment. This result may be because they have severe atopic asthma, and Th2-High endotype, and for that reason they are undergoing more differentiated treatments, such as biological treatments.

Keywords: Asthma, Therapeutics, Machinne Learning, Prediction

References:

1. Asthma GI for. Global Strategy for Asthma Management and Prevention, 2023 [Internet]. Global Strategy for Asthma Management and Prevention; 2023. Disponível em: www.ginasthma.org
2. Dharmage SC, Perret JL, Custovic A. Epidemiology of asthma in children and adults. Vol. 7, Frontiers in Pediatrics. Frontiers Media S.A.; 2019.

2.11. Comparison of accuracy in 2D and 3D templates in total knee arthroplasty

Comparison of accuracy in 2D and 3D templates in total knee arthroplasty

Tânia Viana[1,2,*], Brígida Mónica Faria[1,2]

1. ESS – Polytechnic of Porto, Porto, Portugal; 2. LIACC – Artificial Intelligence and Computer Science Lab (member of LASI), Porto, Portugal.

* Corresponding author email: taniaviana1998@gmail.com

Introduction: Total knee arthroplasty replaces damaged components with prostheses, often used to treat severe osteoarthritis or joint injuries. Advances such as digital templating improve preoperative planning, reducing errors and costs, but accuracy in the choice of implants is crucial to the success of the intervention. This study will compare the accuracy between three-dimensional (3D) and two-dimensional (2D) templating in total knee arthroplasty (TKA), highlighting the expected superiority of 3D templating in accuracy, despite higher costs, compared to 2D templating, which offers a more cost-effective, albeit less accurate, approach. Objectives: The aim of this study is to evaluate the accuracy of three-dimensional (3D) models compared to two-dimensional (2D) models in the context of total knee arthroplasty (TKA) surgery. **Objectives:** The aim of this study is to evaluate the accuracy of three-dimensional (3D)

models compared to two-dimensional (2D) models in the context of total knee arthroplasty (TKA) surgery. **Methods:** This study will be conducted using PeekMed software. Planning will be carried out by orthopedic physicians or accredited personnel in both dimensions. Statistical techniques will be applied to assess accuracy and correlation between clinical variables such as age, gender and height. Statistical analysis may require different approaches depending on the sample size and data distribution. For 2D data, it will be possible to apply parametric tests, such as Student's t-test or analysis of variance (ANOVA). Three-dimensional data may require non-parametric statistical methods, such as the Wilcoxon test or the Kruskal-Wallis test. **Results:** The study to be carried out aims to examine the results of total knee arthroplasty in a sample of 87 participants, 64 of whom are female and 23 male, aged between 47 and 85. It is to be expected that there will be a predominance of two-dimensional data compared to three-dimensional data. This scenario is influenced by factors such as the availability of resources, data collection methods and the complexity associated with generating three-dimensional models. **Conclusion:** The results obtained will be crucial to furthering understanding of the clinical and technical implications of these approaches, and can be presented at medical congresses to foster discussion and progress in orthopaedics. Three-dimensional (3D) models are expected to demonstrate superior accuracy in predicting the actual size of the implant in total knee arthroplasty (TKA) compared to two-dimensional (2D) models, helping to reduce intraoperative errors. Despite the costs associated with 3D models, their possible high accuracy makes them a valuable option for surgical planning of TKA, while 2D models offer a more economical but less accurate approach.

Keywords: Total knee arthroplasty, digital templating, precision.

References:

1. Ioshitake, B.C.A.F, Mendes, E.D, Rossi, F.M, Rodrigues, A.D.C. (2016). Reabilitação de pacientes submetidos à artroplastia total de joelho: revisão de literatura. Revista da faculdade de ciências médicas de Sorocaba v.18 (nº1)
2. Ettinger. M, Claassen, L, Paes, P, Calliess, T. (2016) 2D versus 3D templating in total knee arthroplasty, The Knee v.23 (nº1)
3. PeekMed Blog Orthopedics and heathcare technology (2021) PeekMed® Raises $3 Million in Series A round to continue its expansion

2.12. **Improving genetic diagnostics for patients with negative exome sequencing using AI-powered genome analysis**

Improving genetic diagnostics for patients with negative exome sequencing using AI- powered genome analysis

Joana Ribeiro[1,2*], Brígida Mónica Faria[1,3], Pedro Pinto[2], Joaquim Sá[4]

1. ESS – Polytechnic of Porto, Porto, Portugal; **2**. Quilaban, Sintra, Portugal; **3**. LIACC – Artificial Intelligence and Computer Science Lab (member of LASI), Porto, Portugal.**4**. Unilabs Genetics, Porto, Portugal.

*Corresponding author email: 10220100@ess.ipp.pt

Introduction: Next-generation sequencing (NGS) can be performed using several different platforms. Detecting DNA alterations that affect human health is now possible because of NGS technologies [1]. Despite the advancements in NGS, challenges persist, particularly in cases where traditional exome sequencing yields negative results. Sequencing of the entire genome provides global information about exons and introns, which can reveal the regulatory components of genes, such as promoters, enhancers, and intronic regulators, and structural variants, like copy number variants, inversions, and translocations [2],[3]. As with any technology, DNA sequencing has its limitations. Despite these limitations, DNA sequencing technology has revolutionised our understanding of cellular physiology in health and disease [1]. Although these steps do not directly involve bioinformatics per se, they may have downstream consequences on the bioinformatics algorithms used. **Objectives:** This study focuses on the analysis and performance of genome sequencing in patients with negative exome results. Leveraging AI technology, our study explores both non-coding and coding regions of the genome, aiming to unlock hidden insights. **Methods:** For this project, the data were collected from patients with a negative result for Whole Exome sequencing (WES) and who were resequenced for the Whole Genome Sequencing (WGS) approach. After, the raw data was treated by Emedgene (AI- algorithm-based software) and afterwards the results between techniques were compared and statistically treated using SPSS software. **Results:** The results were based on 16 selected cases with a negative exome diagnosis but with a diagnosis result from other complementary techniques. The first analysis appreciation shows that at least in 80% of the cases, the diagnosis could be assumed with WGS using this specific AI- algorithm. **Conclusions:** The implications of AI genome analysis, showcase the potential to provide more accurate and conclusive genetic diagnoses, improving patient care and treatment decisions. Bridging the gap left by traditional exome sequencing offers a promising avenue for precision medicine and personalized healthcare.

Keywords: Artificial Intelligence; Genetics; Whole-genome sequencing; Bioinformatics.

References:

1. O. Akintunde, T. Tucker, and V. J. Carabetta, "The evolution of next-generation sequencing technologies," ArXiv, p. arXiv:2305.08724v1, May 2023. D. A. Wheeler et al., "The complete genome of an individual by massively parallel DNA sequencing," Nature, vol. 452, no. 7189, pp. 872–876, Apr. 2008, doi: 10.1038/nature06884.
2. J. Chou, T. K. Ohsumi, and R. S. Geha, "Use of whole exome and genome sequencing in the identification of genetic causes of primary immunodeficiencies," Current Opinion in Allergy & Clinical Immunology, vol. 12, no. 6,
3. pp. 623–628, Dec. 2012, doi: 10.1097/ACI.0b013e3283588ca6.
4. N. B. Larson, A. L. Oberg, A. A. Adjei, and L. Wang, "A Clinician's Guide to Bioinformatics for Next- Generation Sequencing," Journal of Thoracic Oncology, vol. 18, no. 2, pp. 143–157, Feb. 2023, doi: 10.1016/j.jtho.2022.11.006.
5. K. F. Oakeson, J. M. Wagner, M. Mendenhall, A. Rohrwasser, and R. Atkinson-Dunn, "Bioinformatic Analyses of Whole-Genome Sequence Data in a Public Health Laboratory," Emerg. Infect. Dis., vol. 23, no. 9, pp. 1441–1445, Sep. 2017, doi: 10.3201/eid2309.170416.
6. A. M. Kanzi et al., "Next Generation Sequencing and Bioinformatics Analysis of Family Genetic Inheritance," Front. Genet., vol. 11, p. 544162, Oct. 2020, doi: 10.3389/fgene.2020.544162.
7. E. L. Van Dijk, H. Auger, Y. Jaszczyszyn, and C. Thermes, "Ten years of next-generation sequencing technology," Trends in Genetics, vol. 30, no. 9, pp. 418–426, Sep. 2014, doi: 10.1016/j.tig.2014.07.001.
8. H.-G. Klein, P. Bauer, and T. Hambuch, "Whole genome sequencing (WGS), whole exome sequencing (WES) and clinical exome sequencing (CES) in patient care," LaboratoriumsMedizin, vol. 38, no. 4, pp. 221–230, Jul. 2014, doi: 10.1515/labmed-2014-0025.

3. Posters

The poster session included a variety of research topics in bioinformatics and biostatistics. Advancements in imaging analysis, including the development of a CellProfiler pipeline for adipocyte differentiation and the application of machine learning for breast cancer tumor classification, were highlighted. Clinical and genomic insights were also presented, including a preliminary analysis of MASK-Air data on allergic rhinitis and its impact on productivity, as well as in-silico modeling of the ataxin-3 protein network relevant to Spinocerebellar Ataxia type 3. Genomic research explored sequence alignment algorithms for KRAS mutations and the evaluation of the REVEL score across diverse contexts.

All posters were available in digital format throughout the event, ensuring continuous access for participants.

3.1. Development of a CellProfiler pipeline to evaluate adipocyte differentiation

Development of a CellProfiler pipeline to evaluate adipocyte differentiation

João Andrade[1,2], Sílvia Torres[1], Pedro Coelho[1,2*]

1. Center for Translational Health and Medical Biotechnology Research (TBIO)/Health Research Network (RISE-Health), ESS, Polytechnic of Porto, Porto, Portugal; 2. ESS, Polytechnic of Porto, Porto, Portugal

* Corresponding author email: pvc@ess.ipp.pt

Introduction: Obesity is a complex chronic disease characterized by excessive body fat accumulation, with increasingly prevalence worldwide, burdening individuals and healthcare systems, thus urgent research is needed. (1,2) Adipocytes, the major cellular component of adipose tissue, are cells vastly used by the scientific community for *in vitro* studies of obesity. (3) Oil red O (ORO) staining and quantification is widely used for intracellular lipid staining and adipogenesis evaluation. Modern microscopy and image analysis software like CellProfiler enable efficient, high-throughput cellular image analysis, improving biological understanding and overcoming manual microscopy processing limitations. (4) **Objectives:** The present work aimed to develop an *in silico* image-based method to evaluated lipid accumulation along the differentiation and adipogenesis of adipocytes. **Methods:** Briefly, 3T3-L1 preadipocytes were differentiated with a cocktail of insulin (10 μg/mL), dexamethasone (1 μM) and 3- isobutyl-1-methylxanthine (0.25 mM) and maintained in culture for 12 days. Brightfield contrast phase images, before and after ORO staining, were captured every two days.

Lipid-droplet accumulation was evaluated by both CellProfiler analysis and ORO quantification.

Results: Throughout differentiation, 3T3-L1 cells exhibited adipocyte- like morphological changes, with increasing lipid accumulation, detected by ORO staining. CellProfiler automated image analysis was comparable to ORO staining quantification, both detecting, approximately after day 4, the presence and accumulation of lipid droplets. **Conclusions:** The results showed that along differentiation of 3T3-L1 cells into mature adipocytes, CellProfiler evaluation of lipid accumulation provided similar results as ORO staining. Altogether, automated *in silico* image-based protocols can be used to investigate adipogenic differentiation *in vitro*, overcoming the demanding conventional quantitative methods.

Keywords: obesity, adipocytes, CellProfiler, image analysis

References:

1. Seidell JC, Halberstadt J. The global burden of obesity and the challenges of prevention. Ann Nutr Metab. 2015;66:7–12.
2. Caballero B. Humans against Obesity: Who Will Win? Adv Nutr. 2019;10:S4–9. Jo J, Gavrilova O, Pack S, Jou W, Mullen S, Sumner AE, et al. Hypertrophy and/or Hyperplasia: Dynamics of Adipose Tissue Growth. PLoS Comput Biol. março de 2009;5(3):e1000324.
3. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler : image analysis software for identifying and quantifying cell phenotypes. 2006;7(10).

3.2. Machine Learning in Tumor Classification in Breast Cancer

Machine Learning in Tumor Classification in Breast Cancer

Ana Sofia Lima[1], Carolina Coutinho[1], Raquel Machado[1*], Alexandra Alves Oliveira[1,2], Brígida Mónica Faria[1,2]

1. ESS, Polytechnic Institute of Porto, 4200-072 Porto, Portugal; **2.** Artificial Intelligence and Computer Science Laboratory, 4150-181 Porto, Portugal

* Corresponding author email: 10231153@ess.ipp.pt

Introduction: Breast cancer is the primary cause of mortality among women worldwide (1). Discernible patterns can be found within the disease, presenting an opportunity for the application of machine learning (ML), garnering effective results in screening and diagnosis. **Objectives:** Different ML algorithms were tested – Decision Tree, Deep Learning (DL), k-Nearest Neighbors (k-NN) and Naïve Bayes - to construct a predictive model allowing the early classification of a breast tumor as benign or malignant, avoiding the need to proceed to a more invasive technique. **Methods:** The ML models were constructed and applied to a database of 201 individuals with breast cancer and descriptive attributes (e.g. age, tumor size, presence of invasive nodes) (2) by using RapidMiner Studio. The evaluation of the models was done by analyzing their accuracy, true negative (TNR) and true positive rates (TPR), their ROC (Receiver Operating Characteristic) curves and AUC (Area Under Curve). **Results:** During a first exploratory phase, fours clusters were detected: smaller tumor sizes, younger patients, and a benign

diagnosis; older age, bigger tumor sizes and a malignant diagnosis; and two more with the opposite characteristics. These characteristics were later found to be important factors in the construction of the Decision Tree. When comparing the models accuracy, the best model was Naïve Bayes (91.04%), followed by the Decision Tree (90.55%), DL (90.02%) and k-NN (86.32%). There is a statistically significant difference between the performances of every model ($p<0.05$) except between the DL and the Decision Tree models. Naïve Bayes presented the highest TPR (98.21%) while DL presented the highest TNR (83.15%). The Decision Tree model presented the highest AUC (0.976), followed by Naïve Bayes (0.961). **Conclusions:** The Decision Tree model best achieved our goal by having the highest AUC which denotes an exceptional sensitivity rate, surpassing Naïve Bayes while maintaining a similar accuracy and TNR.

Keywords: machine learning, predictive models, breast cancer

References:

1. World Health Organization. Global breast cancer initiative implementation framework: assessing, strengthening and scaling-up of services for the early detection and management of breast cancer: executive summary. Geneva; 2023. Available from: https://www.who.int/publications/i/item/9789240067134
2. Eteng I, Bisong E, Fagbola T, Ibrahim M, Udosen J, Akpotuzor S. UCTH Breast Cancer Dataset. Mendeley Data: V2; 2023. Available from: https://data.mendeley.com/datasets/63fpbc9cm4/2 doi: 10.17632/63fpbc9cm4.2

3.3. Allergic rhinitis and work productivity: preliminary analysis of data from the MASK-Air application

Allergic Rhinitis and Work Productivity: Preliminary Analysis of Data from the MASK-air Application

Laura Ferreira[1], Bernardo Sousa Pinto[2], Sandra Maria Alves[1,3], Rita Amaral[1,2]

1.ESS, Polytechnic Institute of Porto, Portugal; **2.**Faculty of Medicine, University of Porto & CINTESIS@RISE, Porto, Portugal;**3.**Centre for Health Studies and Research of the University of Coimbra/Centre for Innovative Biomedicine and Biotechnology (CEISUC/CIBB), Coimbra, Portugal.

Introduction: Allergic rhinitis is a health condition more prevalent in developed countries that can impact the activities and quality of life of affected individuals[1]. Although its impact on work productivity is recognized[2], there is still a need for a more detailed understanding and quantification. This cross-sectional observational study investigates the relationship between allergic rhinitis and work productivity, using data from the MASK-air mobile designed for monitoring allergic rhinitis and related respiratory conditions[3]. **Objective**: To investigate the association between the severity of allergic rhinitis symptoms and the impact on work productivity. **Methods**: Data was collected through the MASK-air mobile application[4,5] that

contains demographic, environmental and symptom variables on a daily basis, with users providing information on a scale of 0 to 100 each day. A sample of 1000 random observations of users from 30 countries, recorded between May 2015 and December 2023 was analysed. Participants were selected based on specifics criteria, including a minimum age of 15 or 16 (depending on the digital consent age in each country) and self-reported diagnosis of allergic rhinitis. Descriptive statistics and the Spearman correlation coefficient[6] between symptoms and impact on productivity were calculated. **Results:** The sample showed a balanced distribution between sexes, with 435 individuals identified as female (53.5%) and 378 individuals as male (46.5%). The mean age of participants was 41.41 ± 14.50 years. The data included participants from various countries; the most frequent was from Mexico with 141 participants (17.3%), followed by Lithuania with 91 participants (11.9%), and Germany with 79 participants (9.7%). Regarding comorbidities, 535
participants (65.6%) reported having conjunctivitis, and 310 participants (38.1%) reported being asthmatic. Additionally, 200 participants (20%) used immunotherapy. A strong positive correlation was observed between work impact and the severities of global allergic symptoms (ρ_s= 0.82, p < 0.0001) and nasal symptoms (ρ_s= 0.77, p <0.0001); and a moderate correlation was observed between work impact and the severities of ocular symptoms (ρ_s= 0.69, p < 0.0001) and asthma (ρ_s= 0.48, p< 0.0001). **Conclusion:** This study offers an initial understanding of how symptoms of allergic rhinitis affect work productivity. Identifying other associated factors will allow targeting health interventions and policies to improve the well-being and performance of workers affected by this condition.

Keywords: Allergic rhinitis, work productivity, mobile health

References:

1. Baiardini I, Braido F, Tarantini F, Porcu A, Bonini S, Bousquet PJ, et al. ARIA-suggested drugs for allergic rhinitis: what impact on quality of life? A GA2LEN review. Allergy. 2008 Jun;63(6):660–9. https://doi.org/10.1111/j.1398-9995.2008.01649.x
2. Bousquet J, Bewick M, Arnavielhe S, Mathieu-Dupas E, Murray R, Bedbrook A, et al. Work productivity in rhinitis using cell phones: The MASK pilot study. Allergy. 2017 Oct;72(10):1475–84. https://doi.org/10.1111/all.13177 3.Passalacqua G, Durham SR. Allergic Rhinitis and its Impact on Asthma update: Allergen immunotherapy. J Allergy Clin Immunol. 2007 Apr;119(4):881–91.
3. Sousa-Pinto B, Schünemann HJ, Sá-Sousa A, Rafael José Vieira, Amaral R, Antó JM, et al. Comparison of rhinitis treatments using MASK-air® data and considering the minimal important difference. Allergy. 2022 Jun 13;77(10): 3002–14.
4. Vieira RJ, Pham-Thi N, Anto JM, Czarlewski W, Sá-Sousa A, Amaral R, et al. Academic Productivity of Young People With Allergic Rhinitis: A MASK-air Study. J Allergy Clin Immunol Pract. 2022 Nov;10(11):3008–17.e4. https://pubmed.ncbi.nlm.nih.gov/35998876/
5. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012 Sep;24 (3):69–71.

3.4. *In-silico* prediction of the complete ataxin-3 protein network relevant for Spinocerebellar Ataxia type 3 (SCA3)

In-silico prediction of the complete ataxin-3 protein network relevant for Spinocerebellar Ataxia type 3 (SCA3)

Paulo Batista[1,2,*], Jorge Vieira[1,3], Cristina P. Vieira[1,3], Brígida Mónica Faria [1,3]

1. I3S Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto Portugal; ESS Escola Superior de Saúde do Politécnico do Porto, Porto, Portugal; **2.** IBMC Instituto de Biologia Molecular e CelularI3S, Porto, Portugal; **3.** LIACC – Laboratório de Inteligência Artificial e Ciência de Computadores

* Corresponding author email: 10220893@ess.ipp.pt

Introduction: Spinocerebellar ataxia type 3, also known as Machado Joseph disease (SCA3/MJD), is the most common inherited ataxia worldwide and is caused by a pathogenic expansion of the polyglutamine (polyQ) tract, located at the C-terminal region of the ataxin-3 protein (1). The polyQ region is involved in the stabilization of protein- protein interactions (PPIs). Abnormal polyQ expansion results in structural changes of the ataxin-3 (2,3), implying different accessibility at specific interacting residues, needed for the normal protein activity. PolyQ proteins have large protein networks. Mapping of PPIs has been performed using high-throughput methods, that are known to produce false interactions (4). Therefore, the use of multiple interactomes comparisons (conserved interactions between pairs of proteins which have interacting homologs in another organism, as well as proteomic data from cell lines, patients, mutants expressing a human protein, and cross-species genetic screens (modifier screens), available at EvoPPI3 (5)), together with in-silico analyses, can be used to support PPIs, as well as identify novel interactors. **Objectives:** In this work we will: 1- characterize ataxin-3 network (validating the proteins identified in main databases, as well as identify new putative interactors); 2- identifying the interactors that behave differently in the presence of an expanded polyQ using different 3D structure prediction methods and protein docking methods. **Methods:** Using EvoPPI3 and protein expression in tissues that matter to SCA3 for PPI retrieval and validation, as well as identification of new interactors. *In-silico* approaches for predicting protein binding differences between wildtype and expanded ataxin-3 forms will be performed, using different *a)* 3D protein structure predictions (namely ITASSER (6), AlphaFold (7), and D-ITASSER (8)) and *b)* protein docking methodologies (such as HADDOCK (9) and ClustPro (10). **Results:** Using EvoPPI3, there are 422 ataxin-3 interactors in human main databases. From this, 250 proteins have been previously studied. Of the remaining 172 proteins, 158 have been reported from proteomic analyses of human cell lines and ataxin-3 patients (*H. sapiens* polyQ_22 database), and these could be true interactors. 28 proteins are in common when considering the polyQ, *Mus musculus* interlogs and *Danio rerio* interlogs, and these could be novel interactors to study. From the 158, 73 proteins bind more to the expanded form of ataxin-3 using AlphaFold, to confirm these results we used ITASSER, where we obtained 46 of the 73 that bind more to

the expanded form. **Conclusion:** This study contributes significantly to understanding SCA3 pathology by delineating a network of ataxin-3 interactors and analysing their behaviour in the presence of an expanded polyQ stretch.

Keywords: ataxin-3, SCA3, polyQ protein-protein interactions, *in-silico* methodology

References:

1. McLoughlin HS, Moore LR, Paulson HL. Pathogenesis of SCA3 and implications for other polyglutamine diseases. Neurobiol Dis [Internet]. 2020 Feb 1 [cited 2024 Jan 17];134. Available from: https://pubmed.ncbi.nlm.nih.gov/31669734/
2. Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell [Internet]. 2006 May 19 [cited 2024 Jan 17];125(4):801–14. Available from: https://pubmed.ncbi.nlm.nih.gov/16713569/
3. Rocha S, Vieira J, Vázquez N, López-Fernández H, Fdez-Riverola F, Reboiro-Jato M, et al. ATXN1 N-terminal region explains the binding differences of wild-type and expanded forms. BMC Med Genomics [Internet]. 2019 Oct 26 [cited 2024 Jan 17];12(1):1–14. Available from: https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0594-4
4. Sousa e Silva R, Sousa AD, Vieira J, Vieira CP. The Josephin domain (JD) containing proteins are predicted to bind to the same interactors: Implications for spinocerebellar ataxia type 3 (SCA3) studies using Drosophila melanogaster mutants. Front Mol Neurosci. 2023 Mar 15;16:1140719.
5. Sousa A, Rocha S, Vieira J, Reboiro-Jato M, López-Fernández H, Vieira CP. On the identification of potential novel therapeutic targets for spinocerebellar ataxia type 1 (SCA1) neurodegenerative disease using EvoPPI3. J Integr Bioinform [Internet]. 2023 Jun 1 [cited 2024 Jan 17];20(2). Available from: https://pubmed.ncbi.nlm.nih.gov/36848492/
6. Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, et al. I-TASSER-MTD: a deep-learning- based platform for multi-domain protein structure and function prediction. [cited 2024 Apr 20]; Available from: https://doi.org/10.1038/s41596-022-00728-0
7. David A, Islam S, Tankhilevich E, Sternberg MJE. The AlphaFold Database of Protein Structures: A Biologist's Guide. J Mol Biol [Internet]. 2022 Jan 30 [cited 2024 Jan 17];434(2). Available from: https://pubmed.ncbi.nlm.nih.gov/34757056/
8. D-I-TASSER: deep learning-based protein structure prediction [Internet]. [cited 2024 Jan 19]. Available from: https://zhanggroup.org/D-I-TASSER/
9. Dominguez C, Boelens R, Bonvin AMJJ. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. J Am Chem Soc [Internet]. 2003 Feb 15 [cited 2024 Jan 17];125(7):1731–7. Available from: https://pubmed.ncbi.nlm.nih.gov/12580598/
10. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, et al. The ClusPro web server for protein-protein docking. Nat Protoc [Internet]. 2017 Feb 1 [cited 2024 Jan 21];12(2):255–78. Available from: https://pubmed.ncbi.nlm.nih.gov/28079879/

Sequence Alignment: Comparative Analysis of Algorithms in KRAS Genetic Mutations

Ana Rita Pereira[1,2], Carlos Lopes[1,3], Catarina Oliveira[1,4], Gonçalo Pereira[1,5], Rui Moreira*[1,6], Brígida Mónica Faria[1,7]

1. ESS – Polytechnic of Porto, Portugal, **2.** Luz Arrabida Hospital, Vila Nova de Gaia, Portugal; **3.** Armed Forces Hospital – Porto Campus, Portugal, **4.** Alto Ave Local Health Unit, EPE – Senhora da Oliveira Hospital, Guimarães, Portugal **5.** Local Health Unit Santo António – Hospital Magalhães Lemos, Porto, Portugal **6.** Local Health Unit São João, EPE – Nossa Senhora da Conceição Hospital, Valongo, Portugal; **7**. LIACC – Artificial Intelligence and Computer Science Laboratory (member of LASI), Porto, Portugal

*Corresponding author email: 10230714@ess.ipp.pt

Introduction: For this study a program was developed in Python on biological sequence alignment, considering the application of algorithms in the genetic analysis of Kristen rat sarcoma viral oncogene (KRAS) and its main mutations associated with cancer. The KRAS gene, like other genes in the same family, is responsible for encoding proteins that regulate cell proliferation, differentiation, and apoptosis. The algorithms employed include the Needleman-Wunsch algorithm as well as the Smith-Waterman algorithm and the Basic Local Alignment Search Tool (BLAST). The algorithms for multiple sequence alignment help to understand the function, evolution, and variability of biological sequences, significantly contributing to advances in genomics and proteomics. **Objectives:** The objectives of this study are to apply algorithms for the effective alignment of biological sequences, compare the non-mutated KRAS sequence with principal mutations associated with cancer development, delineate and justify the selection of the algorithms used, assess their computational complexity, and facilitate 3D visualization of the sequences. **Methods:** Development of a program – **BioAlign** – in Python, with various functions including upload and visualization of sequences; use of different algorithms for global and local alignment; BLAST search; algorithms complexity analysis; obtaining the nucleotides positions; obtaining subsequence and its position; phylogenetic analysis; histogram visualization of sequence length and 3D structure visualization. **Results:** The program is capable of analyzing and comparing the provided sequences using both local and global algorithms. The execution time among the three main algorithms differs, with the BLAST algorithm notably slower in returning results. This fact may be due to several factors, such as the complexity of the algorithm itself, the internet speed, and the response time of the NCBI website. **Conclusions:** The development of the **BioAlign** program indeed allows to address the proposed objectives. Furthermore, the completion of this project has enhanced proficiency in utilizing the Python programming language, demonstrating significant skill development.

Keywords: sequence alignment; algorithm; KRAS.

References:

1. Macedo, D. Alinhamento de sequências biológicas na Bioinformática. Centro Universitário Anhanguera de Niterói; UNIAN Niterói; Rio de Janeiro; 2014.
2. Mariano, DCB (org.) et al. BIOINFO - Revista Brasileira de Bioinformática e Biologia Computacional. 1ª Edição. Volume. 1. ISBN: 978-6-599-275326. Lagoa Santa: Alfahelix,; 2021. DOI: 10.51780/978-6-599-275326
3. Alinhamento global (ulisboa.pt) – Consultado em: 08/04/2024
4. National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov – Consultado em: 09/04/2024
5. Peter M. K. Westcott, Minh D. To. The genetics and biology of KRAS in lung cancer. Chin J Cancer. 32(2):63-70; 2013. doi:10.5732/cjc.012.10098
6. Zanatto, R. M., Santos, G., Oliveira, J. C., Pracucho, E. M., Nunes, A. J. F., Lopes-Filho, G. J., & Saad, S. S. Impact of kras mutations in clinical features in colorectal cancer. ABCD. São Paulo: Arquivos Brasileiros de Cirurgia Digestiva; 33; e1524; 2020.
7. Lopes, R. M. D. S. C. Marcadores moleculares do adenocarcinoma do pulmão. 2012.

3.6. Assessing the Utility of the REVEL Score: A Comprehensive Evaluation Across Diverse Genomic and Clinical Contexts

Assessing the Utility of the REVEL Score: A Comprehensive Evaluation Across Diverse Genomic and Clinical Contexts

Inês Ribeiro[1,2*], Maria Abreu[1,3*], Marta Leão[1*], Miguel Abreu[1*], Brígida Mónica Faria[1,4]

1. ESS – Polytechnic of Porto; **2**. ULS Gaia/Espinho; **3**. ULS Santo António; **4**. LIACC – Artificial Intelligence and Computer Science (member of LASI)

*10190338@ess.ipp.pt,10230709@ess.ipp.pt,10230711@ess.ipp.pt,10230712@ess.ipp.pt

Introduction: Interpreting germline variant pathogenicity is challenging, even with increased access to genomic data and in silico prediction tools. The REVEL score, an ensemble method combining 13 prediction tools, has become a key resource for classifying missense variants. This study evaluates REVEL's accuracy using gnomAD data, focusing on three aspects: its agreement with ClinVar classifications, its reliability with variants of moderate-to-high prevalence in gnomAD 4.0 (which are generally benign), and its effectiveness across gene pathogenicity mechanisms, such as gain of function and loss of function. This analysis will determine REVEL's utility in diverse clinical settings. **Methods:** It was optimized data processing by selecting 20 genes from the OMIM-morbid database, representing a variety of disorders and disease mechanisms. To test the accuracy of REVEL, it was selected genes with varying features, focusing on pathogenicity mechanisms (such as gain of function, loss of function, or dominant negative), inheritance patterns (autosomal dominant, autosomal recessive, or X-linked), and disorder frequencies. This approach allowed us to evaluate REVEL's

performance across diverse gene characteristics and clinical scenarios.It was mapped each gene's REVEL score to its gnomAD frequency, ClinVar classification, and canonical transcript position, and accuracy was tested using Python and Biopython. **Results/Discussion:** Our preliminary analysis showed that the REVEL score performed well for variants with medium–to–high prevalence in gnomAD. REVEL scores were generally consistent with ClinVar classifications, with high accuracy across most gene type, but some care should be taken upon analysing Clinvar classification, as some may have used REVEL or some of its components during interpretation. The tool was effective regardless of pathogenicity mechanisms, inheritance patterns, or disorder frequencies, suggesting broad utility in genomic analysis.

Keywords: Score REVEL, gnomAD, missense variants, ClinVar classifications

References:

1. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, Biesecker LG. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. PLoS Genet. 2016 Sep 23;12(9):e1005854. doi: 10.1371/journal.pgen.1005854. PMID: 27666373; PMCID: PMC5033142.
2. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 2015 Apr 1;24(8):2125–37. doi: 10.1093/hmg/ddu733. Epub 2015 Jan 2. PMID: 25552646; PMCID: PMC4355007.
3. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062–D1067. doi: 10.1093/nar/gkx1153. PMID: 29165669; PMCID: PMC5753288.

4. Invited Talks: MBBAS Alumni

The SBBAH also counted on the participation of three Alumni from the MBBAS program, who delivered invited talks during a session of the Symposium:

Sara Silva (Master's graduate in 2019) presented "Exploring the Functional Activity of Astrocytes and Their Crosstalk with Neurons." Her research provided valuable insights into the functional roles of astrocytes and their interactions with neurons, advancing our understanding of neural communication and brain function.

Cláudio Correia (Master's graduate in 2021) delivered a talk titled "A Saúde no Digital – Análise de Dados de Saúde Pública," where he examined the use of digital tools in the analysis of public health data, highlighting the importance of data analysis in driving informed decisions and improving public health outcomes.

Beatriz Torres (Master's graduate in 2023) presented "Application of Machine Learning Techniques for a Recommendation Assistant in Pharmacy." Her work focused on the contributions for a recommendation assistant using machine learning algorithms, aiming to improve the decision-making processes in the pharmaceutical industry.

These Alumni presentations emphasized diverse approaches to solving complex problems across fields such as neuroscience, public health, and machine learning.

5. Companies and Research Labs

The contributions of companies and research labs emphasized the synergy between academic research and industry.

The involvement of companies highlighted the fact that research also takes place beyond academia, emphasising that the effective translation of scientific knowledge is greatly enhanced when academia and enterprises work together. The presentations included **Glintt Global**, represented by Catarina Faria, who discussed the company's strategies in tech talent acquisition to drive innovation in health-related fields. Francisco Sousa, from **B-Simple Healthcare Solutions**, presented insights on data analysis in critical health areas, focusing on the development of new tools to support clinical and administrative decision-making. Ricardo Pereira, from **OmniumAI**, presented CibusAI, highlighting the role of artificial intelligence in food innovation and reformulation.

Academic contributions were equally impactful. The research labs at the Escola Superior de Saúde of the Instituto Politécnico do Porto, specialize in the analysis and interpretation of health data, with the goal of advancing clinical decision-making, patient outcomes, and public health initiatives. Their contributions demonstrate the role of academic research in health and illustrate how biostatistics and bioinformatics drive progress in the health sciences. Mónica Vieira introduced the **TBIO** Center for Translational Health and Medical Biotechnology Research (E2S-P.PORTO and TBIO@RISE), emphasizing its role in bridging translational research and medical biotechnology. Agostinho Cruz, representing **CISA** (a health and environment research hub of LAQV/REQUIMTE at E2S/IPP), presented its role as a center for advancing interdisciplinary research. Additionally, Paulo Veloso Gomes highlighted the work of the Rehabilitation Research Center (**CIR/LabRP**), illustrating its active contributions to health sciences through rehabilitation research. Beatriz Sousa, from the Faculty of Medicine at the University of Lisbon, presented **iSTARS**, which focuses on informatics and statistical tools to enhance research success.

This proceedings book gathers the contributions presented at the scientific event SBBAH 2024, held in Hospital das Forças Armadas – Polo Porto (HFAR), on 3 May 2024.
The book was edited and published in 2025 by ESS | P. PORTO Edições, and obtained its ISBN registration that same year.