IWBBIO-2023

PROGRAM & ABSTRACTS

11th-14th JULY, 2023
Gran Canaria (SPAIN)

IWBBIO-2023 Program

Tuesday, July 11th, 2023	
18:30-20:00	**REGISTRATION DESK** ***(start at 18:30h but it is open during all the conference)***
18:30-20:00	**Upload the presentations to the room's computer (in case you haven't sent them by email).**

NOTES:

- All **Sessions A** will be held in Hotel Lopesan Villa del Conde Resort. They are **face-to-face sessions**, and they will also be shared on-line by Zoom. The **plenary lectures** are in **Session A.**
- All **Sessions B** will be held on-line (virtual) using Zoom.
- Oral Presentation: **17 minutes** (including the questions). Depending on whether there are absent speakers, times may be adjusted.
- **Poster** authors are requested to place their posters on the panels before the start of the poster session (e.g. morning posters can be placed before 10 o'clock, before the coffee break, and afternoon posters before 16:00. They can be placed on any of the panels). The morning posters should be removed at 14:30 and the afternoon posters at 19:45. It is recommended to use **A0 size** and large fonts.

Session A: Located in the last floor of the main building.

Wednesday, July 12, 2023		
8:30	**REGISTRATION DESK** ***(start at 8:30h but it is open during all the conference)***	
	All **Sessions A: Oral face-to-face sessions**. All **Sessions B:** Oral (will be held on-line by Zoom)	
9:00-10:10	**Session A.1: E-Health and Computational Support for Clinical Decisions**	**Session B.1: Biomedical Engineering**
10:10- 11:00	**Session A.2: Biomarker Identification**	**Session B.2: Analysis of Molecular Dynamics Data in Proteomics**
11:00-11:40	**COFFEE BREAK**	
11:45-12:45	**Session A.P1: Opening & Plenary Lecture.** **Dr. Sotirios Kiokias** **European Research Executive Agency. Research Programme Manager. Chemistry Panel leader, Marie Sklodowska Curie Actions (MSCA)**	
12:45-14:00	**Session A.3: Biomedical Engineering (Part I)**	**Session B.3: Computational Support for Clinical Decisions**
14:00-16:00	**REST BREAK**	
16:00-16:55	**Session A.4: Recent Advances in COVID 19**	**Session B.4: Image Visualization and Signal Analysis**
17:00-18:30	**Session A.5: Analysis of Molecular Dynamics Data in Proteomics**	**Session B.5: Machine Learning in Bioinformatics**

Thursday, July 13th, 2023		
8:30	**REGISTRATION DESK** ***(start at 8:30h but it is open during all the conference)***	
	All **Sessions A: Oral face-to-face sessions**. All **Sessions B:** Oral (will be held on-line by Zoom)	
9:00-9:45	**Session A.6: Biomedical Engineering (Part II)**	**Session B.6: Biomarker Identification**
9:45- 10:45	**Session A.7: High-throughput genomics: bioinformatic tools and medical applications**	**Session B.7: New Computational Approaches in Biomedicine**
10:45-11:15	**COFFEE BREAK**	
11:15-12:15	**Session A.P2: Plenary Lecture.** **Prof. Ashok Mulchandani** **Professor at University of California. Pioneer Faculty of Bourns College of Engineering. Distinguished Professor. Bionanotechnology and Biosensors Group**	
12:15-14:00	**Session A.8: Image Visualization and Signal Analysis**	**Session B.8: New Advances in Bioinformatics and Biomedicine (Part I)**
14:00-16:00	**REST BREAK**	
16:00-16:55	**Session A.9: Sensor-Based Ambient Assisted Living Systems and Medical Applications**	**Session B.9: New Advances in Bioinformatics and Biomedicine (Part II)**
17:00-18:20	**Session A.10: Biomedical Computing**	
18:20 - 19:30	**Session A.11: POSTER SESSION (A)**	
20:30	**GALA DINNER** **Hotel Lopesan Baobab 5*** **(15 minutes walking from Hotel Lopesan Villa del Conde Resort)**	

<table>
<tr><th colspan="3">Friday, July 14th, 2023</th></tr>
<tr><td rowspan="2">8:30</td><td colspan="2">REGISTRATION DESK
(start at 8:30h but it is open during all the conference)</td></tr>
<tr><td colspan="2">All Sessions A: Oral <u>face-to-face sessions</u>.
All Sessions B: Oral (will be held on-line by Zoom)</td></tr>
<tr><td>9:00-10:10</td><td>Session A.12: Machine learning in Bioinformatics and NGS</td><td rowspan="2">Session B.10: New Advances in Bioinformatics and Biomedicine (Part III)</td></tr>
<tr><td>10:10- 11:00</td><td>Session A.13: Feature Selection, Extraction, and Data Mining in Bioinformatics: Approaches, Methods and Adaptations</td></tr>
<tr><td>10:00 -14:30</td><td colspan="2">Session A.14: POSTER SESSION (B)</td></tr>
<tr><td>11:00-11:45</td><td colspan="2">COFFEE BREAK</td></tr>
<tr><td>11:45-12:45</td><td colspan="2">Session A.P3: Plenary Lecture.
Prof. Alfredo Vellido
Universitat Politècnica de Catalunya.
Full Professor. Department of Computer Science (CS).</td></tr>
<tr><td>12:45-14:00</td><td>Session A.15: Advanced in Bioinformatics</td><td></td></tr>
</table>

IWBBIO 2023 PROGRAM

Wednesday, July 12, 2023

(9:00-10:10) Session A.1: E-Health and Computational Support for Clinical Decisions

Chairman: Dr. Francesco Fioranelli, Dr. Hanna Piotrzkowska Wróblewska and Dr.Nuno Rosa

Predicting and detecting coronary heart disease in patients using machine learning method **(Ref: *1758*)**

Michal Wos, Bartłomiej Kiczek and Bartłomiej Drop

Deep Learning for Parkinson's Disease Severity Stage Prediction using a New Dataset **(Ref: *7587*)**

Zainab Maalej, Fahmi Ben Rejab and Kaouther Nouira

Improved Long-term Forecasting of Emergency Department Arrivals with LSTM-based Networks **(Ref: *9307*)**

Carolina Miranda-Garcia, Alberto Garces-Jimenez, Jose Manuel Gomez-Pulido and Helena Hernandez-Martinez

Measurement Of Acute Pain In The Pediatric Emergency Department Through Automatic Detection Of Behavioral Parameters: A Pilot Study **(Ref: *9753*)**

Letizia Bergamasco, Marco Gavelli, Carla Fadda, Emilia Parodi, Claudia Bondone and Emanuele Castagno

(9:00-10:10) Session B.1: Biomedical Engineering

Chairman: Dr. Francisco Ortuño

Motion control of a robotic lumbar spine model **(Ref: *1997*)**

Thuanne Paixão, Ana Beatriz Alvarez, Ruben Florez and Facundo Palomino-Quispe

Improving Foetal Health Monitoring: A Review of the Latest Developments and Future Directions **(Ref: *4049*)**

Restuning Widiasih, Hasballah Zakaria and Siti Saidah Nasution

Transparent Machine Learning Algrithms for Explainable AI on Motor fMRI Data **(Ref: *6449*)**

Jose Diogo Marques dos Santos, David Machado and Manuel Fortunato

Analyzing dose parameters of Radiation Therapy Treatment Planning and Estimation of Second Cancer Risks **(Ref: *8632*)**

Irine Khomeriki, Lily Petriashvili, Maia Topeshashvili and Tamar Lominadze

(10:10- 11:00) Session A.2: Biomarker Identification

Chairman: Dr. Antonio Mucherino and Dr. Andrzej Kloczkowski

Predicting Cancer Stage from Circulating microRNA: A Comparative Analysis of Machine Learning Algorithms **(Ref: *5616*)**

Sören Richard Stahlschmidt, Benjamin Ulfenborg and Jane Synnergren

Osteopontin overexpression synergistically interacts with Aurora kinases overexpression, and is associated with tumor progression, early tumor recurrence, and poor prognosis in hepatocellular carcinoma **(Ref: *7684*)**

Zhong-Zhe Lin, I-Lun Tsai and Kuan-Yu Chen

Gait Asymmetry Evaluation using FMCW Radar in Daily Life Environments **(Ref: *9807*)**

Shahzad Ahmed, Yudam Seo and Sung Ho Cho

(10:10- 11:20) Session B.2: Analysis of Molecular Dynamics Data in Proteomics

Chairman: Dr. Alfredo Vellido and Dr. Caroline Konig

Recognition of conformational states of a G Protein-Coupled Receptor from molecular dynamic simulations using sampling techniques **(Ref: *2884*)**

Mario Alberto Gutiérrez Mondragón, Caroline König and Alfredo Vellido Alcacena

Inter-helical residue contact prediction in -helical Transmembrane proteins using structural features **(Ref: *6505*)**

Aman Sawhney, Jiefu Li and Li Liao

Structural Analysis of RNA-Binding Protein EWSR1 Involved in Ewing's Sarcoma through Domain Assembly and Conformational Molecular Dynamics Studies **(Ref: *8336*)**

Saba Shahzadi, Mubashir Hassan and Andrzej Kloczkowski

Degree-normalization improves random-walk-based embedding accuracy in PPI graphs **(Ref: *6924*)**

Luca Cappelletti, Stefano Taverni, Tommaso Fontana, Marcin P. Joachimiak, Justin Reese, Peter Robinson, Elena Casiraghi and Giorgio Valentini

---

(11:45-12:45) Opening Ceremony. Plenary Talk: Dr. Sotirios Kiokias

European Research Executive Agency. Research Programme Manager. Chemistry Panel leader, Marie Sklodowska Curie Actions (MSCA).

Title of the presentation: EU funding opportunities of Horizon Europe. Marie Sklodowska Curie Actions (MSCA)

(12:45-14:00) Session A.3: Biomedical Engineering (Part I)

Chairman: Dr. Andrzej Swierniak and Dr. Amita Barik

Evolutionary games in modeling cancer metastases **(Ref: *2735*)**

Andrzej Swierniak, Katarzyna Hajdowska and Damian Borys

Antimicrobial ceramic materials for biomedical implants **(Ref: *8536*)**

Julietta V. Rau

Investigation of inclusion for localised characteristics from medical imaging datasets genotype-phenotype associations **(Ref: *6568*)**

Gabrielle Dagasso, Matthias Wilms and Nils Forkert

Data Augmentation Techniques for Improving Biomedical Name Entity Recognition in Low-Resource Settings **(Ref: *6833*)**

Yiling Cao, Zhongguang Zheng and Lu Fang

Portable MRI System Based on the Gradient-Free Imaging Technique **(Ref: *7317*)**
Boguslaw Tomanek, Aaron Purchase, Christoper Sedlock and Jonathan Sharp

(12:45-14:00) Session B.3: Computational Support for Clinical Decisions

Chairman: Dr. Francisco Ortuño

Validation of Height-for-Age and BMI-for-Age Z-scores Assessment using Android-based Mobile Apps **(Ref: *1742*)**
Valerii Erkudov, Sergey Lytaev, Kenjabek Rozumbetov, Andrey Pugovkin, Azat Matchanov and Sergey Rogozin

Systematic comparison of advanced network analysis and visualization of lipidomics data **(Ref: *3767*)**
Jana Schwarzerová, Dominika Olešová, Aleš Kvasnička, David Friedecký, Margaret Varga, Valentine Provazník and Wolfram Weckwerth

Comparison of image processing and classification methods for a better diet decision-making **(Ref: *4206*)**
Maryam Abbasi, Pedro Martins and Filipe Cardoso

A Machine Learning approach to predict brain abnormalities in preterm infants using clinical data **(Ref: *5481*)**
Arantxa Ortega Leon, Roa'A Khaled, María Inmaculada Rodríguez García, Daniel Urda and Ignacio Turias

Ethical dilemmas, mental health, artificial intelligence and LLM based chatbots **(Ref: *5739*)**
Johana Cabrera, Soledad Loyola, Irene Magaña and Rodrigo Rojas

(16:00-16:55) Session A.4: Recent Advances in COVID 19

Chairman: Dr. Xiaohong WANG and Dr. Vicente Martínez

Application of custom approaches for target enrichment of SARS-CoV-2 in nasopharyngeal swab for whole genome sequencing **(Ref: *3517*)**
Anna Gladkikh, Ekaterina Klyuchnikova, Valerya Sbarzaglia, Dmitrii Polev and Vladimir Dedkov

Physiological polyphosphate: a new molecular paradigm in biomedical applications for human therapy **(Ref: *7038*)**
Prof. Dr. Werner E. G. Müller and Prof. Dr. Xiaohong Wang

Quantitative EEG Findings in Outpatients with Psychosomatic Manifestations after COVID-19 **(Ref: *8024*)**

Sergey Lytaev, Nikita Kipaytkov and Tatayna Navoenko

(16:00-16:55) Session B.4: Image Visualization and Signal Analysis

Chairman: Dr. Aleksandr Sinitca and Dr. Lily Petriaschvili

Digital Breast Tomosynthesis reconstruction techniques in healthcare systems: A review **(Ref: *5473*)**

Imane Samiry, Ilhame Ait Lbachir, Imane Daoudi, Saida Tallal and Sayouti Adil

BCAnalyzer: A semi-automated tool for the rapid quantification of cell monolayer from microscopic images in scratch assay **(Ref: *7437*)**

Aleksandr Sinitca, Airat Kayumov, Pavel Zelenikhin, Andrey Porfiriev, Dmitrii Kaplun and Mikhail Bogachev

Breast Cancer Histologic Grade Identification by Graph Neural Network Embeddings **(Ref: *8322*)**

Salvatore Calderaro, Giosuè Lo Bosco, Filippo Vella and Riccardo Rizzo

---

(17:00-18:30) Session A.5: Analysis of Molecular Dynamics Data in Proteomics

Chairman: Dr. Alfredo Vellido and Dr. Caroline Konig

The Coherent Multi-Representation Problem with Applications in Structural Biology **(Ref: *1151*)**

Antonio Mucherino

Prediction of Functional Effects of Protein Amino Acid Mutations **(Ref: *4683*)**

Óscar Álvarez-Machancoses, Eshel Faraggi, Enrique Deandrés-Galiana, Juan Fernández-Martínez and Andrzej Kloczkowski

Computational study of conformational changes in intrinsically disordered regions during protein-protein complex formation **(Ref: *4804*)**

Amita Barik, Madhabendra Mohon Kar and Prachi Bhargava

Conserved Water Networks Identification in Proteins Using Density Clustering Approaches on Positional and Orientational Data **(Ref: *1884*)**
Urban Bren, Marko Jukic, Jelena Tosovic and Domagoj Fijan

(17:00-18:30) Session B.5: Machine Learning in Bioinformatics

Chairman: Dr. Francisco Ortuño

Stochastic model of infection with the SARS–COV–2 virus in a small group of individuals indoors **(Ref: *1100*)**
Igor Derevich and Anastasiia Panova

An efficient algorithm for detecting mutually exclusive patterns across multiple sets of genomic mutations **(Ref: *2555*)**
Siyu He, Jiayin Wang, Zhongmeng Zhao and Xuanping Zhang

Exploring machine learning algorithms and protein language models strategies to develop enzyme classification systems **(Ref: *3798*)**
Diego Fernández, Álvaro Olivera-Nappa, Roberto Uribe-Paredes and David Medina

Relation Predictions in Comorbid Disease Centric Knowledge Graph Using Heterogeneous GNN Models **(Ref: *4664*)**
Saikat Biswas, Koushiki Dasgupta Chaudhuri, Pabitra Mitra and Krothapalli Sreenivasa Rao

GPU Cloud architectures for bioinformatic applications **(Ref: *8985*)**
Antonio Maciá-Lillo, Tamai Rodríguez, Higinio Mora, Antonio Jimeno-Morenilla and Jose-Luís Sánchez-Romero

Cyclical learning rates (CLR's) for improving training accuracies and lowering computational cost **(Ref: *9523*)**
Shrikant Pawar, Aditya Stanam, Anand Narayanan and Rushikesh Chopade

Thursday, July 13th, 2023

(9:00-9:45) Session A.6: Biomedical Engineering (Part II)

Chairman: Dr. Werner E. G. Müller and Dr.Vesna Rastija

Annotation-free Identification of Potential Synteny Anchors **(Ref: *7451*)**
Karl-Kristian Kaether, Steffen Lemke and Peter F. Stadler

Clinical text classification in Cancer Real-World Data in Spanish **(Ref: *3696*)**
Francisco J. Moreno-Barea, Héctor Mesa, Nuria Ribelles, Emilio Alba and Jose M. Jerez

(9:00-9:50) Session B.6: Biomarker Identification

Chairman: Dr. Francisco Ortuño

The pathogenetic significance of miR-143 in atherosclerosis development **(Ref: *491*)**
Mikhail Lopatin, Maria Vulf, Maria Bograya, Anastasia Tynterova and Larisa Litvinova

Significance of using of Liquid Biopsy for Guiding Therapy Decision in Cancer **(Ref: *770*)**
Omayma Mazouji, Hicham Mansour and Abdelhak Ouhajjou

Complex Network and Artificial Intelligence combined approach to investigate Autism spectrum disorder through gene expression data **(Ref: *6114*)**
Antonio Lacalamita, Alfonso Monaco, Nicola Amoroso, Loredana Bellantuono, Alessandro Fania, Ester Pantaleo, Sabina Tangaro and Roberto Bellotti

(9:45- 10:45) Session A.7: High-throughput Genomics: Bioinformatic Tools and Medical Applications

Chairman: Dr. Jose Luis Oliveira

A 20-year journey of tracing the development of web catalogues for rare diseases **(Ref: *4883*)**
João Almeida and José Luis Oliveira

Unsupervised investigation of information captured in pathway activity score in scRNA-Seq analysis **(Ref: *5952*)**
Kamila Szumala, Joanna Polanska and Joanna Zyla

Meta-analysis of gene activity (MAGA) contributions and correlation with gene expression, through GAGAM. **(Ref: *6258*)**
Lorenzo Martini, Roberta Bardini, Alessandro Savino and Stefano Di Carlo

---

(9:55- 11:15) Session B.7: New Computational Approaches in Biomedicine

---

Chairman: Dr. Francisco Ortuño

Deep Learning for automatic electroencephalographic signals classification **(Ref: *3046*)**
Nadia N. Sánchez-Pozo, Samuel Lascano-Rivera, Francisco J. Montalvo-Marquez and Dalia Y. Ortiz-Reinoso

Pharmacoinformatics Analysis of Drug Leads for Alzheimer's Disease from FDA-Approved Dataset through Drug Repos itioning Studies **(Ref: *4285*)**
Mubashir Hassan, Saba Shahzadi and Andrzej Kloczkowski

MetaLLM: Residue-wise Metal ion Prediction Using Deep Transformer Model **(Ref: *6035*)**
Fairuz Shadmani Shishir, Bishnu Sarker, Farzana Rahman and Sumaiya Shomaji

Predicting Papillary Renal Cell Carcinoma Prognosis Using Integrative Analysis of Histopathological Images and Genomic Data **(Ref: *7286*)**
Shaira Kee, Michael Aaron Sy, Samuel Border, Nicholas Lucarelli, Akshita Gupta, Pinaki Sarder, Marvin Masalunga and Myles Joshua Tan

Assessing Temporal Stability of Heart Rate Variability Features for Predicting Adverse Cardiovascular Events in Hypertensive Patients **(Ref: *9515*)**
José María López Belinchón, Miguel Á. López and Raúl Alcaraz

(11:15-12:15) Plenary Talk: Prof. Ashok Mulchandani

Professor at University of California. Pioneer Faculty of Bourns College of Engineering. Distinguished Professor.

Bionanotechnology and Biosensors Group.

Title of the presentation: Towards ASSURED Diagnostics Using Paper-Microfluidic Integrated Chemiresistor Biosensor Arrays

---

(12:15-14:10) Session A.8: Image Visualization and Signal Analysis

Chairman: Dr. Veska Gancheva

Correlation of tumor density and computed tomography response criteria to time to tumor progression of patients with advanced hepatocellular carcinoma receiving anti-angiogenic therapy in clinical trials **(Ref: *527*)**
Zhong-Zhe Lin and Po-Chin Liang

Color Hippocampus Image Segmentation using Quantum Inspired Firefly Algorithm and Merging of Channel-wise Optimums **(Ref: *8290*)**
Alokeparna Choudhury, Sourav Samanta, Sanjoy Pratihar and Oishila Bandyopadhyay

A pilot study of neuroaesthetics based on the analysis of electroencephalographic connectivity networks in the visualization of different dance choreography styles **(Ref: *8602*)**
Almudena González, Jose Melendez and Julian J González

The promise of deep learning-assisted multimodality medical image analysis (extended talk) **(Ref: *5233*)**
Habib Zaidi

(12:15-14:00) Session B.8: New Advances in Bioinformatics and Biomedicine (Part I)

Chairman: Dr. Francisco Ortuño

Genomic characterization and phylogenomic analyses of the Beta variant of SARS-CoV-2 circulated in Pakistan **(Ref: *272*)**
Nazia Fiaz, Imran Zahoor, Saima Mahad, Tahir Yaqub and Atia Basheer

Whole tumor area estimation in incremental brain MRI using dilation- and erosion-based binary morphing **(Ref: *573*)**
Orcan Alpar and Ondrej Krejcar

Trimeric receptor-binding domain of SARS-CoV-2 acts as a potent inhibitor of ACE2 receptormediated viral entry **(Ref: *691*)**
Parameswaran Ramakrishnan

AI(Aritificial Intelligence) Determination of Functional Ambulatory Category of Stroke Patients **(Ref: *788*)**
Shi-Uk Lee, Seong-Ho Jang, Youngkook Kim, Jeong-Hyun Kim and Chang Han Lee

BodyFlow: A library for human pose estimation and activity recognition **(Ref: *3869*)**
Irene López-Bosque, Carlos Marañes-Nueno, Ana Caren Hernández-Ruiz, Rocío Aznar-Gimeno, Pilar Salvo-Ibañez, María de La Vega Rodrigálvarez-Chamarro, David Abadía-Gallego and Rafael del-Hoyo-Alonso

---

(16:00-16:55) Session A.9: Sensor-Based Ambient Assisted Living Systems and Medical Applications

---

Chairman: Dr. Ivan Miguel Serrano Pires and Dr. Paulo Jorge Simões Coelho

Smart Wearables Data Collection and Analysis for Medical Applications: A Preliminary Approach for Functional Reach Test **(Ref: *5477*)**
João Duarte, Luis Francisco, Ivan Miguel Pires and Paulo Coelho

Using Digital Biomarkers for Objective Assessment of Perfusionists' Workload and Acute Stress during Cardiac Surgery **(Ref: *8802*)**
Roger Daglius Dias, Lauren Kennedy-Metz, Rithy Srey, Geoffrey Rance, Mahdi Ebnali, David Arney, Matthew Gombolay and Marco Zenati

---

(16:00-18:30) Session B.9: New Advances in Bioinformatics and Biomedicine (Part II)

Chairman: Dr. Marco Mesiti and Dr. Irene López Bosque

Novel gene signature for Bladder Cancer Stage identifcation **(Ref: *1027*)**
Iñaki Hulsman, Luis Javier Herrera, Francisco Ortuño and Ignacio Rojas

Three-dimensional representation and visualization of high-grade and low-grade glioma by Nakagami imaging **(Ref: *1799*)**
Orcan Alpar and Ondrej Krejcar

Role of parallel processing in brain magnetic resonance imaging **(Ref: *2503*)**
Ayca Kirimtat and Ondrej Krejcar

Identification of Novel Anti-inflammatory Agents by Molecular Docking Studies **(Ref: *2557*)**

Sadaf Naeem, Atia Shaheen and Sadaf Naeem

Hybradization of Emperical Mode Decomposition and Machine Learning for Categorization of Cardiac Diseases **(Ref: *2666*)**

Saeed Mian Qaisar

Quality of Life Analysis of Patients with Dermatological Problems: Teledermatology Versus Face-to-Face Dermatology **(Ref: *2785*)**

Antonio López-Villegas, Remedios López Liria and Maria Angeles Valverde-Martinez

A System Biology and Bioinformatics approach to determine the molecular signature, core ontologies, functional pathways, drug compounds in between Stress and Type 2 Diabetes **(Ref: *4092*)**

Md. Abul Basar, Md. Rakibul Hasan, Bikash Kumar Paul, Khairul Alam Shadhin and Md. Sarwar Mollah

EMCNN: Fine-Grained Emotion Recognition based on PPG using Multi-scale Convolutional Neural Network **(Ref: *4218*)**

Jiyang Han and Hui Yang

MODULATION OF CYP1A1 AND CYP1B1 GENE EXPRESSION IN 3-METHYLCHOLANTHRENE-INDUCED PROSTATE CANCER BY AFRICAN PEAR FRUIT (DACRYODES EDULIS) IN WISTAR RATS. **(Ref: *4885*)**

David Omisore

A Meta-Graph for the Construction of an RNA-centered Knowledge Graph **(Ref: *7008*)**

Emanuele Cavalleri, Sara Bonfitto, Alberto Cabri, Jessica Gliozzo, Paolo Perlasca, Mauricio Soto-Gomez, Gabriella Trucco, Elena Casiraghi, Giorgio Valentini and Marco Mesiti

---

(17:00-18:20) Session A.10: Biomedical Computing

Chairman: Dr. Roderick Melnik

Speeding up simulations for radiotherapy research by means of machine learning **(Ref: *2323*)**

Luis Javier Herrera, Ignacio Rojas, Francisco Carrillo Pérez, Alberto Guillen, Isabel Fernández and Carmen Ovejero

Preliminary Results of Using the Tangram Meta-Heuristic for Virtual Screening in Drug Discovery **(Ref: *6563*)**
N.C. Cruz, S. Puertas-Martin, J.L. Redondo and P.M. Ortigosa

Modelling of Anti-Amyloid-Beta Therapy for Alzheimer's Disease **(Ref: *8476*)**
Swadesh Pal and Roderick Melnik

(18:20-19:30) Session A.11: POSTER SESSION (A)

Chairman: Dr. Luis Javier Herrera

Bayesian Molecular Dating Analyses Combined with Mutational Profiling Suggest an Independent Origin and Evolution of SARS-CoV-2 Omicron BA.1 and BA.2 Sub-Lineages **(Ref: *734*)**
Naveen Kumar, Rahul Kaushik, Ashutosh Singh, Vladimir N. Uversky, Kam Y. J. Zhang, Upasana Sahu, Sandeep Bhatia and Aniket Sanyal

Comparison of VCFs generated from different software in the evaluation of variants in genes responsible for rare thrombophilic conditions **(Ref: *1236*)**
Radek Vrtel, Petr Vrtel and Radek Vodicka

Developing a workflow to gain insights from scRNA-seq data for drug development : a case-study with human pancreatic adenocarcinoma **(Ref: *1267*)**
Octavio Morante-Palacios, Ester Gil Vazquez, Oleg Deryagin and Cecilia Klein

Predictive model for Age-Related Macular Degeneration Response to Anti-VEGF treatment **(Ref: *1279*)**
Álvaro Pérez Sala, Rafael Pelaez, Ana Isabel Oca Lázaro, Ángela Villanueva Martinez and Ignacio M. Larráyoz Roldán

Targeted Next Generation Sequencing of a custom capture panel to target sequence 112 cancer related genes in breast cancer tumours ERBB2 positive from Lleida (Spain) **(Ref: *2083*)**
Ana Velasco, Izaskun Urdanibia and Serafín Morales

Oral e-Health Monitoring Platform - a platform for sharing oral health data **(Ref: *2784*)**
Mónica Fernandes, Eduardo Esteves, Diogo Duarte, Marlene Barros, Nuno Rosa and André Correia

Putative quadruplex forming sequences in plants: role, occurrence and possible interaction partners **(Ref: *2789*)**
Adriana Volná, Martin Bartas and Jiří Červeň

Training Strategies for Covid-19 Severity Classification using Machine Learning and Heart Rate Variability Metrics **(Ref: *3356*)**

Daniel Pordeus Menezes, Pedro Ribeiro, Laíla Zacarias, João Madeiro, João Marques, Pedro Rodrigues, Camila Leite, Manoel Neto, Arnaldo Peixoto Jr and Adriel de Oliveira

Medical X-ray Image Classification Method Based on Convolutional Neural Networks **(Ref: *4486*)**

Veska Gancheva, Tsviatko Jongov and Ivaylo Georgiev

WATER DYNAMICS IN CHEESE BY MEANS OF NUCLEAR MAGNETIC RESONANCE RELAXOMETRY **(Ref: *4821*)**

Monika Małkowska-Kowalczyk, Justyna Żulewska, Adriana Lobacz, Maciej Maciejczyk and Danuta Kruk

Bioinformatics for Transcription Factor Discovery at a Primarily Undergraduate Institution **(Ref: *5108*)**

Michael Van Dyke and John Barrows

Deep learning systems for the classification of cardiac pathologies using ECG signals **(Ref: *5307*)**

Olga Valenzuela, Ignacio Rojas-Valenzuela, Fernando Rojas, Juan Carlos De la Cruz and Peter Gloesekoetter

The effectiveness of quarantine in viral and bacterial epidemics: new evidence provided by the Covid-19 pandemic **(Ref: *5421*)**

Andreu Martínez-Hernández and Vicente Martínez

The dark side of NCBI: Annotation artifacts across the RefSeq database **(Ref: *5608*)**

Martin Bartas, Jiří Červeň, Adriana Volná and Petr Pečinka

Recent Advances in Discovery of New Tyrosine Ki-nase Inhibitors Using Computational Methods **(Ref: *6932*)**

Vesna Rastija and Maja Molnar

Preliminary Study on the Identification of Diseases by Electrocardiography Sensors' Data **(Ref: *7246*)**

Rui João Pinto, Pedro Miguel Silva, Rui Pedro Duarte, Francisco Alexandre Marinho, António Jorge Gouveia, Norberto Jorge Gonçalves, Paulo Jorge Coelho, Eftim Zdravevski, Petre Lameski, Nuno Garcia and Ivan Miguel Pires

Assessing bioinformatic tools for de novo assembly of nanopore sequencing data from human whole-genomes **(Ref: *7881*)**
Adrián Muñoz-Barrera, Víctor García-Olivares, Luis A. Rubio-Rodríguez, David Jáspez, José M. Lorenzo-Salazar, Rafaela González-Montelongo and Carlos Flores

Modelling the survival kinetics of Salmonella spp. on the surface of ripened raw milk cheese during storage at different temperatures **(Ref: *8515*)**
Adriana Łobacz and Justyna Zulewska

Eukaryotic topoisomerases of type IIA: cytoplasmic proteins? **(Ref: *9048*)**
Jiří Červeň, Martin Bartas and Petr Pečinka

Investigating the Dynamics of Cancer Evolution: Variant Allele Frequency Patterns and Model Limitations **(Ref: *9284*)**
Paweł Kuś and Marek Kimmel

Target NGS data analysis identifies the haplotype in LRRK2 gene as a potential risk factor for endemic parkinsonism **(Ref: *9396*)**
Radek Vodicka, Kristyna Kolarikova, Petr Kanovsky and Radek Vrtel

The use of H-Scan ultrasound imaging to assess the re-sponse of breast cancer patients to neoadjuvant chemo-therapy **(Ref: *9863*)**
Hanna Piotrzkowska Wróblewska, Katarzyna Dobruch Sobczak, Ziemowit Klimonda, Piotr Karwat and Jerzy Litniewski

Friday, July 14th, 2023

(9:00-10:10) Session A.12: Machine learning in Bioinformatics and NGS

Chairman: Dr. Silvia BOTTINI and Dr. Radek Vrtel

Machine learning combining multi-omics data and network algorithms identifies Adrenocortical carcinoma prognostic biomarkers **(Ref: *247*)**

Roberto Martin-Hernandez

Competitive Analysis of 5 S AND 16 S bacterial RNA in the Phylogeny of bacteria **(Ref: *2158*)**

Michael Sadovsky and Yulia Ovchinnikova

A Platform for the Study of Drug Interactions and Adverse Effects Prediction **(Ref: *5387*)**

Diogo Mendes and Rui Camacho

Enabling real-time analysis of nanopore 16S rRNA sequencing data with NanoRTax **(Ref: *9662*)**

Héctor Rodríguez Pérez, Laura Ciuffreda and Carlos Flores

(9:00-11:00) Session B.10: New Advances in Bioinformatics and Biomedicine (Part III)

Chairman: Dr. Luis Javier Herrera and Dr. Francisco Ortuño

Effect of the primary macroelements in the susceptibility to diseases: A Review **(Ref: *4962*)**

Grethel Lázara Sieiro Miranda, Alberto Nicolás González Marrero, Eida Luisa Rodríguez Lema and Mérida Rodríguez Regal

Analysis of the organization of the structure of human genes depending on their tissue specificity **(Ref: *5581*)**

Sergey Slyusarev and Olga Lyangasova

Uterine Cervix and Corpus Cancers Characterization through Gene Expression Analysis Using the KnowSeq Tool **(Ref: *5863*)**

Lucía Almorox, Luis Javier Herrera, Francisco Ortuño and Ignacio Rojas

Predicting the Risk of Recurrence and Prognosis in Patients with Hepatocellular Carcinoma **(Ref: *6459*)**

Chi-Chang Chang

A guide and mini-review on the performance evaluation metrics in binary segmentation of magnetic resonance images **(Ref: *6799*)**

Ayca Kirimtat and Ondrej Krejcar

ITRAQ-based proteomic analysis reveals potential osteogenesis-promoted role of ATM in strontium-incorporated titanium implant **(Ref: *7436*)**

Yuzi Xu, Yangbo Xu and Fuming He

Selection Process of Phytochemicals and Efficacy of Thymol, Eugenol and Calcium Ferulate on Heterotrophic Plate Count Bacteria in Water **(Ref: *7653*)**

Humayun Wali and Muhammad Zafar

Principaux facteurs de prise de poids pendant le confinement COVID-19 dans une population Marocaine : une étude transversale **(Ref: *8452*)**

Nourelyakine Lakhdar, Driss Lamri and Moulay Ouahidi

(10:10- 11:00) Session A.13: Feature Selection, Extraction, and Data Mining in Bioinformatics: Approaches, Methods and Adaptations

---

Leveraging latent representation in metabolomics **(Ref: *331*)**

Evariste Njomgue-Fotso, Justine Labory, Youssef Boulaimen and Silvia Bottini

Entropy approach of processing for fish acoustic telemetry data to detect atypical behavior during welfare evaluation **(Ref: *2926*)**

Jan Urban

Determining HPV Status in Patients with Oropharyngeal Cancer from 3D CT Images Using Radiomics: Effect of Sampling Methods **(Ref: *4314*)**

Kubra Sarac and Albert Guvenis

(10:00- 14:30) Session A.14: POSTER SESSION (B)

Chairman: Dr. Ignacio Rojas

HEALTH CARE AND DISEASES **(Ref: *2379*)**

Aditi Katake

The Utilization of Click Reactions in Understanding the Structure and Function of IgE in Allergic Reactions **(Ref: *3365*)**

Parth Shinde, Yash Saini, Veeky Baths, Manas Mandal and Shounak Bhattacharya

Radar Sensing in Healthcare: Challenges and Achievements in Human Activity Classification Vital Signs Monitoring **(Ref: *5617*)**

Francesco Fioranelli, Ronny Guendel, Nicolas Kruse and Alexander Yarovoy

The effect of biofeedback on learning the wheelie position on manual wheelchair **(Ref: *5730*)**

Antonio Pinti

Bioinformatics approaches to characterize the Monkeypox virus genomes from cases of the mid-2022 outbreak **(Ref: *7471*)**

Adrián Muñoz-Barrera, Laura Ciuffreda, Julia Alcoba-Florez, Luis A. Rubio-Rodríguez, Héctor Rodríguez-Pérez, Helena Gil-Campesino, Diego García-Martínez de Artola, Josmar Salas-Hernández, Julia Rodríguez-Núñez, Antonio Íñigo-Campos, Víctor García-Olivares, Oscar Díez-Gil, Rafaela González-Montelongo, Agustín Valenzuela-Fernández, José M. Lorenzo-Salazar and Carlos Flores

Constructing a Stroke Diagnosis and Prognosis System Based on the BPN Algorithm Using Tc-99m-ECD SPECT images **(Ref: *7485*)**

Jui-Jen Chen, Hung-Nien Chang Chien and Yen-Hsiang Chang

A Platform for the Radiomic Analysis of Brain FDG PET Images: Detecting Alzheimer's disease **(Ref: *9490*)**

Ramin Rasi and Albert Guvenis

Detecting Intra Ventricular Haemorrhage in Preterm Neonates using LSTM Autoencoders **(Ref: *9647*)**

Idris Muniru, Jacomine Grobler and Lizelle Van Wyk

Insulin sensitivity and Insulin Resistance and Diabetes prediction by applying artificial intelligence techniques on genes obtained from expression analysis differential **(Ref: *1590*)**

Jesús María González-Martín, Francisco Rodríguez-Esparragón, Bernardino Clavo, Laura Beatriz Torres-Mata, Sara Estefania Cazorla-Rivero and Estrella Gómez-Bentolila

Agent based modeling of fish shoal behavior **(Ref: *2829*)**

Pavla Urbanova, Ievgen Koliada, Petr Cisar and Milos Zelezny

Evaluation of homogeneity of effervescent tablets containing quercetin and calcium using X-ray microtomography and hyperspectral analysis **(Ref: *4874*)**
Michał Meisner, Piotr Duda, Beata Szulc-Musioł and Beata Sarecka-Hujar

Modeling and Simulation of Multiphase Flow in Integrated Multitrophic Aquaculture Systems with Macroalgae: Application of CFD-DEM **(Ref: *5017*)**
Radomir Filip, Pavla Urbanova, Ingrid Masalo and Stepan Papacek

The dilemma of choosing the best prediction model and feature engineering approach for heterogenous datasets. **(Ref: *5192*)**
Lukasz Piorecki and Joanna Polanska

Identification of InhA-inhibitors interaction fingerprints that affect residence time **(Ref: *6973*)**
Magdalena Ługowska and Marcin Pacholczyk

Optimizing Variant Calling for Human Genome Analysis: A Comprehensive Pipeline Approach **(Ref: *8458*)**
Miguel Pinheiro, Jorge Miguel Silva and José Luis Oliveira

(11:15-12:15) Plenary Talk: Prof. Alfredo Vellido

Universitat Politècnica de Catalunya. Full Professor. Department of Computer Science (CS).

Title of the presentation: AI as a socialite: time to think about its impact in biomedicine

---

(12:45-14:00) Session A.15: Advanced in Bioinformatics

Chairman: Dr. João Paulo Madeiro

Revealing the RBP regulome in hepatocellular carcinoma via consensus GRN inference **(Ref: *1853*)**
Mateusz Garbulowski, Riccardo Mosca, Carlos J. Gallardo-Dodd, Claudia Kutter and Erik L. L. Sonnhammer

GeneCaRNA: A new world of non-coding RNAs for disease decipherment **(Ref: *2118*)**
Doron Lancet

An Algorithm for Pairwise DNA Sequences Alignment **(Ref: *2271*)**
Veska Gancheva and Hristo Stoev

Multiallelic Maximal Perfect Haplotype Blocks with Wildcards via PBWT **(Ref: *4398*)**
Paola Bonizzoni, Gianluca Della Vedova, Yuri Pirola, Raffaella Rizzi and Mattia Sgrò

Abstracts book IWBBIO-2023

Submission index: [247], [272], [331], [491], [527], [573], [691], [734], [770], [788], [1027], [1100], [1151], [1236], [1267], [1279], [1590], [1742], [1758], [1799], [1853], [1884], [1997], [2083], [2118], [2158], [2271], [2323], [2379], [2503], [2555], [2557], [2666], [2735], [2784], [2785], [2789], [2829], [2884], [2926], [3046], [3356], [3365], [3517], [3696], [3767], [3798], [3869], [4049], [4092], [4206], [4218], [4285], [4314], [4398], [4486], [4664], [4683], [4804], [4821], [4874], [4883], [4885], [4962], [5017], [5108], [5192], [5233], [5307], [5387], [5421], [5473], [5477], [5481], [5581], [5608], [5616], [5617], [5730], [5739], [5863], [5952], [6035], [6114], [6258], [6449], [6459], [6505], [6563], [6568], [6799], [6833], [6924], [6932], [6973], [7008], [7038], [7246], [7286], [7317], [7436], [7437], [7451], [7471], [7473], [7485], [7587], [7653], [7684], [7881], [8024], [8290], [8322], [8336], [8452], [8458], [8476], [8515], [8536], [8602], [8632], [8802], [8985], [9048], [9284], [9307], [9396], [9490], [9515], [9523], [9612], [9647], [9662], [9753], [9807], [9863]

[247] *Machine learning combining multi-omics data and network algorithms identifies Adrenocortical carcinoma prognostic biomarkers*

Roberto Martin-Hernandez (Clarivate Analytics).

Background: Adrenocortical carcinoma (ACC) is a rare endocrine cancer with a poor prognosis. The knowledge about ACC pathogenesis is incomplete, and patients have limited options. Identification of molecular drivers and effective biomarkers is required for timely diagnosis of the disease and stratify patients to offer the most beneficial treatments. In this study we demonstrate how machine learning methods integrating multi-omics data, in combination with system biology tools, can contribute to the identification of new prognostic biomarkers for ACC. Methods: ACC gene expression and DNA methylation datasets were downloaded from the Xena Browser (GDC TCGA Adrenocortical Cancer cohort). A highly correlated multi-omics signature discriminating groups of samples was identified with the data integration analysis for biomarker discovery using latent components (DIABLO) method. Additional regulators of the identified signature were discovered using Clarivate CBDD (Computational Biology for Drug Discovery) network propagation and hidden nodes algorithms on a curated network of molecular interactions (Metabase). The discriminative power of the multi-omics signature and their regulators was delineated by training a random forest classifier using 55 samples, with 5 runs of a 10-fold cross-validation approach. The prognostic value of identified biomarkers was further assessed on an external dataset obtained from GEO (GSE49280) using the Kaplan-Meier estimator method. Results: [...] A multi-omics signature including 85 highly correlated features (25 genes, 28 microRNA's and 32 methylation sites) was generated. Systems biology tools identified 17 additional genes regulating the features included in the multi-omics signature. Both sets of features revealed a high power to classify patients from stages I-II and stages III-IV with AUC values of 0.78 and 0.81 respectively, outperforming previously identified prognostic biomarkers. Association of the genes included in the signature with overall survival (OS) data demonstrated that patients with upregulation of HOXC10, HAUS8, S1PR3, UBE2S, down-regulation of IGIP, ZCRB1 and miR-125b-2 showed a statistically significant decrease in OS. Interestingly, HAUS8 and miR-125b-2 were found to be biomarkers of the disease with a high level of validity in dedicated databases. Conclusions: Machine learning and integrative analysis of multi-omics data, in combination with Clarivate CBDD systems biology tools, identified a set of biomarkers with high prognostic value for ACC disease. Multi-omics data is a promising resource for the identification of drivers and new prognostic biomarkers in rare diseases that could be used in clinical practice.

---

[272] *Genomic characterization and phylogenomic analyses of the Beta variant of SARS-CoV-2 circulated in Pakistan*

Nazia Fiaz (University of Veterinary and Animal Sciences, Lahore, Pakistan), Imran Zahoor (University of Veterinary and Animal Sciences, Lahore, Pakistan), Saima Mahad (University of Veterinary and Animal Sciences, Lahore, Pakistan), Tahir Yaqub (University of Veterinary and Animal Sciences, Lahore, Pakistan) and Atia Basheer (University of Veterinary and Animal Sciences, Lahore, Pakistan).

Background: In this study, we performed the genomic characterization and phylogenomic analysis of beta-VOC of SARS-CoV-2 in Pakistan, in the context of the global population of this variant.

Methods and Results: A set of 105 whole-genome sequences of Pakistani samples, retrieved from GISAID, were aligned in MAFFT and used as an input to the Coronapp web application for mutation identification. The genome sequence of

Wuhan/Hu-1/2019 (NC_045512.2), was used as the reference genome. The phylogenetic tree was constructed through the neighbor-joining method by downloading the 800 whole-genome sequences of beta-variant from ten different countries having the largest number of Pakistani diasporas. In total, we detected 389 mutations, out of which 227 were missense mutations. We found NSP3 and Spike as the most variable proteins, with 70 and 54 mutations in Pakistan, respectively. Surprisingly, some missense mutations which are also known as characteristic mutations like T265I, K1655N, K3353R (ORF1a), and S84L (ORF8) and del241/243 (S), and had 92–99% prevalence globally, were not present in the Pakistani population of beta-VOC. Moreover, N501Y and E484K substitution mutations in spike protein and L242(S), S106(NSP6) deletions were found to have 52%, 50%, 49%, and 73.3% prevalence in Pakistan, respectively, whereas globally these mutations were 86%, 85%, 84%, and 91% prevalent. Likewise, three missense mutations S794L(NSP3), G30R(N), and W29L(ORF7b) were found to have just a 12%, 0.4%, and 0.3% prevalence globally, but in Pakistan, these mutations had 67%, 67% and 49% prevalence, respectively. The phylogeny analysis revealed that the majority of the Pakistani samples were clustered together with South African samples, followed by samples from England, Saudi Arabia, Oman, and Italy which suggest the transmission of this variant to Pakistan from these countries.

Conclusions: In this study, we found 22 characteristic mutations of beta-variant, of which 14 were present with a prevalence of 95-90%, whereas some other characteristic mutations with high prevalence (92-99%) like T265I, K1655N, K3353R (ORF1a), S84L (ORF8), and del 241/243 (S) were missing in the Pakistani population of beta-variant.

---

[331] *Comprehensive characterization of the role of FMRP in neuronal physiology using multi-omics integration.*

Silvia Bottini (MDLab - MSI - Université Cote d'Azur).

Fragile X-Syndrome (FXS) represents the most common inherited form of intellectual disability and the first monogenic cause of Autism Spectrum Disorders. In most of cases, this disease results from the absence of expression of the protein FMRP (Fragile X messenger ribonucleoprotein). FMRP is a multifunctional RNA binding protein that affects numerous cellular pathways that vary according to the stage of brain development. To accomplish its functions, FMRP interacts with several thousands of mRNA targets to regulate their metabolism and many protein partners in different sub-cellular compartments. Consequently, although FXS is a monogenic disease, an effective strategy will rely on multi-therapy approaches. The development of these compensatory therapies requires first a broad identification of the cellular functions regulated by FMRP and dysregulated in FXS as well as the evaluation and prioritization of their impact on the aetiology of FXS. In addition, the molecular mechanisms underlying these regulatory functions have to be finely understood. In this context, the FXS research community has produced numerous omics studies to evaluate the multiple perturbations of cellular homeostasis in the context of the FXS. Indeed, omics approaches aim to understand a biological system in its entirety through systematic analyses of its content at the molecular level. While there is a plethora of multi-omics data performed and collected, the integration of those remains challenging.

For this purpose, we developed a novel deep-learning based on Variational AutoEncoders (VAE) approach to analyze horizontally integrated multi-omics datasets, called HIVE (Horizontal Integration analysis using VaE). Briefly, VAE can capture non-linear relationships and encode them in the latent space. We coupled the VAE with a random forest regression (RFR) to allow the explainability of the latent features.

We have conducted an unprecedented literature mining study to collect omics data evaluating multiple perturbations of cellular homeostasis in the context of the FXS. We collected 37 studies split into 7 types of omics ranging from proteomics, transcriptomics, metabolomics, lipidomics etc. We created a horizontally integrated dataset for the mouse model consisting of 51 samples from 2 transcriptome datasets, 3 translatome datasets and one proteomic dataset and 3997 common measured molecules. The application of HIVE to this dataset selected 7 features of the latent space. Then, we retrieved 1227 molecules that have contributed the most to these discriminative features by using the contribution score calculated with the RFR. Next, we checked whether these 7 features contained already known interactors of FMRP and/or targets identified by CLIP-seq studies and/or autism-related genes. By inspecting the literature, we selected 6 lists of relevant genes, namely: 2 regarding FRMP partners, 3 target lists from different CLIP-seq studies and one of the autism-related genes. By performing market basket analysis, we have found that feature 15 is strongly associated mainly with the three CLIP-seq studies, showing that this feature is mainly depicting the most known role of FMRP as an RNA-binding protein. Importantly, feature 37 is associated with autism-related genes, suggesting that this latent feature captures the processes associated with this pathology and is potentially regulated by FRMP. Then, we analyzed the pathway composition of these 7 latent features and we provided a comprehensive list of biological pathways in which FMRP is involved. Consistently, term enrichment analysis yielded “Intellectual disabilities” and “RNA binding” as the most significant terms according to the DisGeNet database and the GO Molecular Function classification respectively. As expected, by querying the database SynGO, an expert-curated resource for synapse function, we observed a significant enrichment of several terms. Interestingly, terms connected with RNA splicing and kinase activity also emerged as very

significant by our analysis. To inspect further the list of pathways we performed network analysis using the STRING database and Cytoscape. By selecting only the molecules contributing to at least two features, we retrieved 453 molecules as drivers of the studied phenotype. We applied stringent criteria to retrieve the associations among these molecules from STRING (confidence score>0.7) and we obtained 499 interactions (edges). Then we applied the Markov clustering algorithm to divide the network into modules. For each module, we performed enrichment analysis to identify interacting nodes involved in significant pathways. We found clusters composed of molecules involved mainly in RNA binding, splicing, amino acids metabolism, axon guidance, and kinase activities. Taken together, we demonstrated that HIVE is a powerful approach to analysing horizontally integrated datasets. The application of HIVE to the context of FXS provided the characterization of the most salient pathways associated with phenotype and shed light on novel roles in which FRMP is involved. Although the preliminary results offered a comprehensive overview of the multitude of roles in which FMRP is involved, the horizontal integration that considers only molecules measured in all omics to integrate is a limitation of the model. The next step will be to develop a late integration strategy in which omics datasets will be integrated at the level of the latent space, in order to consider all the measured molecules in each omic dataset and also to integrate omics datasets that cannot be horizontally joined (metabolomics, lipidomics). In the mid-term, this study combined with experimental validations will allow to propose of biological markers and therapeutic targets for FXS with a high potential of reliability.

---

[491] *The pathogenetic significance of miR-143 in atherosclerosis development*

Mikhail Lopatin (Immanuel Kant Baltic Federal University (IKBFU)), Maria Vulf (Immanuel Kant Baltic Federal University (IKBFU)), Maria Bograya (Immanuel Kant Baltic Federal University (IKBFU)), Anastasia Tynterova (Immanuel Kant Baltic Federal University (IKBFU)) and Larisa Litvinova (Immanuel Kant Baltic Federal University (IKBFU)).

Our pilot studies of blood plasma in patients with comorbidities allow miR-143 to be regarded as a potential biomarker of atherosclerosis expression, but the litera-ture data on the dynamics of miR-143 expression are too controversial. The con-tinuation of the study to verify the results by "wet lab" methods is costly and therefore it is advisable to first determine whether the putative biomarker has pathogenic relevance. The aim of the study was to establish and assess the role of miR-143 in atherosclerosis using a comprehensive bioinformatics analysis, to identify possible pathways of its inclusion in the pathology mechanism. Using open sources, two sets of gene expression data in atherosclerotic plaques were selected, then only differentially expressed genes (DEGs) shared by both datasets were identified, from which a protein-protein interaction (PPI) extended network was then constructed. The next step was network analysis (identification of clus-ters and hub genes within them) and construction of regulatory networks of miRNAs-hub genes. The analysis revealed that miR-143 is one of the central miRNAs whose action may be associated with suppression of atherosclerosis formation through its targeting of several hub genes: TLR2, TNF and LYN. However, another target is ITGB1, whose reduction increases autophagy and ac-tivates the inflammatory response. Based on established topological and function-al characteristics, miR-143 is of interest for further verification as a biomarker and for possible therapeutic applications in atherosclerosis. We also cut through the perspective of the already known effects of miR-143 on the NK-kB pathway, but in the context of atherosclerosis (via TNF and TLR2).

---

[527] *Correlation of tumor density and computed tomography response criteria to time to tumor progression of patients with advanced hepatocellular carcinoma receiving anti-angiogenic therapy in clinical trials*

Zhong-Zhe Lin (Natinaol Taiwan University Cancer Center) and Po-Chin Liang (Natinaol Taiwan University Hospital).

Background: Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 is widely used for tumor response assessment in patients with advanced hepatocellular carcinoma (HCC). This study explored the value of RECIST 1.1, mRECIST, Choi criteria, and various tumor density assessment in predicting time to tumor progression (TTP) of patients with advanced HCC receiving anti-angiogenic systemic therapy in clinical trials. Method: A retrospective blinded analysis of computed tomography (CT) scans was performed by two independent radiologists. The anti-angiogenic systemic therapy included sorafenib+UFUR, thalidomide+UFUR, bevacizumab+capecitabine, and bevacizumab+erlotinib. TTP was calculated according to Kaplan-Meier method. Multivariate analyses were conducted by fitting a Cox proportional hazards model. Results: A total of 39 patients were enrolled for the image analysis. Median TTP was 3.7 months, and median overall survival was 6.1 months. The first radiologist identified 5%, 18%, and 49% objective responses by RECIST 1.1, mRECIST, and Choi criteria, respectively. The objective responses evaluated by another radiologist were 8%, 13%, and 41% (RECIST 1.1, mRECIST, and Choi criteria), respectively. The inter-radiologist agreement rates were 79%, 67%, and 72% by RECIST 1.1, mRECIST, and Choi criteria, respectively. Multivariate analysis revealed sorafenib+UFUR, tumor density (normal liver

as reference), and RECIST 1.1 were independent factors to predict TTP for these patients. Conclusions: RECIST 1.1 is a better predictor for TTP of patients with advanced HCC receiving anti-angiogenic systemic therapy, compared to Choi or mRECIST criteria. Tumor density provides additional predictive information for these patients.

---

[573] *Whole tumor area estimation in incremental brain MRI using dilation- and erosion-based binary morphing*

Orcan Alpar (University of Hradec Kralove) and Ondrej Krejcar (University of Hradec Kralove).

Magnetic resonance imaging (MRI) technology is rapidly advancing and three-dimensional (3D) scanners started to play an important role on diagnosis. However, not every medical center has access to 3D magnetic resonance imaging (MRI) devices; therefore, it is safe to state that the majority of MRI scans are still two-dimensional. According to the setup values adjusted before any scan, there might be consistent gaps between the MRI slices, especially when the increment value exceeds the thickness. The gap causes miscalculation of the lesion volumes and misjudgments when the lesions are reconstructed in three-dimensional space due to excessive interpolation. Therefore, in this paper, we introduce three types of morphing methods, one dilation-based and two erosion-based, and compare them to figure out which one provides better solution for filling up the gaps in incremental brain MRI. Among three types of morphing methods, the highest average dice score coefficient (DSC) is calculated as %91.95, which is obtained by the multiplicative dilation morphing method for HG/0004 set of BraTS 2012.

---

[691] *Trimeric receptor-binding domain of SARS-CoV-2 acts as a potent inhibitor of ACE2 receptormediated viral entry*

Parameswaran Ramakrishnan (Case Western Reserve University).

The COVID-19 pandemic has caused over four million deaths and effective methods to control CoV-2 infection, in addition to vaccines, are needed. The CoV-2 binds to the ACE2 on human cells through the receptor-binding domain (RBD) of the trimeric spike protein. Our modeling studies show that a modified trimeric RBD (tRBD) can interact with three ACE2 receptors, unlike the native spike protein, which binds to only one ACE2. We found that tRBD binds to the ACE2 with 58-fold higher affinity than monomeric RBD (mRBD) and blocks spike-dependent pseudoviral infection over 4-fold more effectively compared to the mRBD. Although mRBD failed to block CoV-2 USA-WA1/2020 infection, tRBD efficiently blocked the true virus infection in plaque assays. We show that tRBD is a potent inhibitor of CoV-2 through both competitive binding to the ACE2 and steric hindrance, and has the potential to emerge as a first-line therapeutic method to control COVID-19. https://doi.org/10.1016/j.isci.2022.104716. The full article is already published.

---

[734] *Bayesian Molecular Dating Analyses Combined with Mutational Profiling Suggest an Independent Origin and Evolution of SARS-CoV-2 Omicron BA.1 and BA.2 Sub-Lineages*

Naveen Kumar (National Institute of High Security Animal Diseases), Rahul Kaushik (RIKEN), Ashutosh Singh (National Institute of High Security Animal Diseases), Vladimir N. Uversky (Department of Molecular Medicine, Morsani College of Medicine, University of South Florida), Kam Y. J. Zhang (RIKEN), Upasana Sahu (National Institute of High Security Animal Diseases), Sandeep Bhatia (National Institute of High Security Animal Diseases) and Aniket Sanyal (National Institute of High Security Animal Diseases).

The ongoing evolution of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) has resulted in the recent emergence of a highly divergent variant of concern (VOC) defined as Omicron or B.1.1.529. This VOC is of particular concern because it has the potential to evade most therapeutic antibodies and has undergone a sustained genetic evolution, resulting in the emergence of five distinct sub-lineages. However, the evolutionary dynamics of the initially identified Omicron BA.1 and BA.2 sub-lineages remain poorly understood. Herein, we combined Bayesian phylogenetic analysis, mutational profiling, and selection pressure analysis to track the virus's genetic changes that drive the early evolutionary dynamics of the Omicron. Based on the Omicron dataset chosen for the improved temporal signals and sampled globally between November 2021 and January 2022, the most recent common ancestor (tMRCA) and substitution rates for BA.1 were estimated to be that of 18 September 2021 (95% highest posterior density (HPD), 4 August-22 October 2021) and 1.435 × 10-3 (95% HPD = 1.021 × 10-3 - 1.869 × 10-3) substitution/site/year, respectively, whereas 3 November 2021 (95% highest posterior density (HPD) 26 September-28 November 2021) and 1.074 × 10-3 (95% HPD = 6.444 × 10-4 - 1.586 × 10-3) substitution/site/year were estimated for the BA.2 sub-lineage. The findings of this study suggest that the Omicron BA.1 and BA.2 sub-lineages originated independently and evolved over time.

Furthermore, we identified multiple sites in the spike protein undergoing continued diversifying selection that may alter the neutralization profile of BA.1. This study sheds light on the ongoing global genomic surveillance and Bayesian molecular dating analyses to better understand the evolutionary dynamics of the virus and, as a result, mitigate the impact of emerging variants on public health.

---

[770] *Significance of using of Liquid Biopsy for Guiding Therapy Decision in Cancer*

Omayma Mazouji (Mohammed first university), Hicham Mansour (Mohammed first university) and Abdelhak Ouhajjou (Al azhar oncology center).

Introduction: Liquid biopsies especially those involving circulating tumor DNA (ctDNA) from plasma gained momentum interest in the area of genomic alterations detection and treatment guidance. Liquid biopsy is rapidly emerging as an important complement tool to tumor biopsies and, in some cases, is considered as a potential alternative approach. The purpose of this study is to evaluate, in real life, the ability of ctDNA to detect genetic alteration in patients, to screen the most popular abnormalities, to detect the mutations associated with treatment resistance and to guide therapy decision making for positive patients. Materials and methods: 82 samples were performed in our laboratory from 71 recruited patients in our center between January 2017 and January 2023, as part of their routine clinical follow-up. 15 various cancers were analyzed (breast cancer, lung cancer, colorectal cancer, pancreatic cancer, sarcoma, kidney cancer, gastric cancer, glioblastoma, adenoid cystic cancer, urothelial cancer, desmoid cancer, cervix cancer, Ewing cancer, adrenocortical cancer and endometrial cancer). Genetic testing was conducted in patients who are metastatic and after failure of all standard treatments. Six gene-panel tests were available and one of them was suggested to each patient, the gene-panel tests are ranging from 11 to 105 genes. For each panel performed, we conducted deep sequencing (>100X) of well-known genes involved in cancer. Sequencings were conducted mainly in blood-ctDNA samples of patients; and when it was possible on the tumor tissue as well. Thus, 51 patients had tests conducted on their blood specimens while, 14 patients had tests carried out on their tissue samples, and 6 patients had both tests performed on their tissue and blood specimens (some patients had more than two tests). Results: Our result indicated that liquid biopsy data are similar to tissue biopsy, in term of library quality control, data throughput generated, sequencing data quality, bioinformatics analysis and genetic alteration detection. By analyzing the results generated, we found that 88% of our patients were eligible for treatment guidance using liquid biopsy (63% with available FDA-approved drug, 10% could join clinical trials for experimental drugs and 15% have to avoid some drugs associated with tumor resistance. Only, 11% patients have no detected mutations and 1% patients with detected mutation but without available drug or open clinical trial. NGS allowed us to detect various genetic alterations; in our cohort 80.5% (66/82) of cases were detected with at least two mutations, 11% (9/82) of cases had no detected mutation, and, 8.5% (7/82) of cases were detected with only one mutation. According to identified alterations, a total of 70 genes were identified with 349 distinct aberrations, as follows; 47% deletions, 30% transitions, 11% transversions, 11% duplications and 1% deletion/insertion. Among mutated genes detected in our cohort we observed that BRCA2, EGFR2, MSH6 and NF1 genes were the most mutated in more than 13 among 15 analyzed cancer types. Conclusion: Our study demonstrates that circulating tumor DNA could be used as a biomarker to manage patients with different types of cancer, detect genomic alterations at frequencies similar to those observed in tumor sequencing identify mutations resistance as well as guide therapy decision.

---

[788] *AI(Aritificial Intelligence) Determination of Functional Ambulatory Category of Stroke Patients*

Shi-Uk Lee (Department of Rehabilitation Medicine, Seoul National University College of Medicine, Boramae Medical Center), Seong-Ho Jang (Department of Rehabilitation Medicine, Hanyang University, College of Medicine), Youngkook Kim (Rbiotech Co., Ltd), Jeong-Hyun Kim (Department of Rehabilitation Medicine, Seoul National University College of Medicine) and Chang Han Lee (Department of Rehabilitation Medicine, Gyeongsang National University College of Medicine, Changwon Hospital).

Background: The assessment of gait function in stroke patients is an important evaluation for predicting ambulation, participation in society and underlying prognosis since it depends on the individual's muscle strength, range of motion, balance and coordination. Most gait analysis is qualitative in nature based on the experiences of clinicians and sensor-based analysis which has temporal, spatial, and economic limitations. FAC(Functional Ambulatory Category) has been widely utilized because it does not need any complicated device. However, it is a subjective evaluation tool. With the help of AI technology, it is possible to perform an objective evaluation in such cases. In this study, we classified gait functions using AI by extracting and analyzing coordinate data acquired from viewing gait performance videos of stroke patients. Method: We collected gait videos (1 million pixels, 30fps, mp4) of normal people and stroke patients from 6 medical

institutions to learn the AI model for gait function classification of stroke. Time series joint coordinate data was extracted from 2D JSON and this was pre-processed (deleting missing value, standard deviation value of each joint). The pre-processed data were randomly divided into a training set (80%) for AI model learning and a test set (20%) for performance verification. The AI model was deep-learned with a train set to classify it into 'Normal,' 'Mild,' 'Moderate,' and 'Severe' according to gait function, and 5-fold cross validation was performed. The accuracy of the model was verified with the test set. The classification of 'Normal,' 'Mild,' 'Moderate,' and 'Severe' was based on the scores of FAC evaluated by physiatrists. Result: The subjects who participated in the study were normal 278 (male:52, female:226), mean age 64.9, stroke 218 (male:134, female:84), mean age 62.2. Stroke type was ischemic in 135 patients and hemorrhagic in 83 patients. Injury region of brain was cortical in 156 patients and subcortical in 35 patients. As a result of verifying the performance of the AI model based on gait video, the accuracy of stroke classification for normal persons was 95.09% and Area Under the Curve (AUC) was 0.94. The functional gait classification accuracy of stroke patients was 83.44% and AUC was 0.95. Conclusion: AI-based gait function classification was confirmed by utilizing video files of gait performance in stroke patients. This tool served as support data for functional gait evaluation and diagnosis performed by physiatrists. In this study, walking video files only were applied without using various types of sensors. The AI model based on image data had an accuracy of over 80%. This approach suggests the potential for developing IT technologies and applications for non-face-to-face care and patient self-monitoring. In future studies, it will be necessary to analyze and learn the patterns of gait motions using time series data of gait videos taken from various angles.

[1027] *Novel gene signature for Bladder Cancer Stage identifcation*

Iñaki Hulsman (Department of Computer Engineering, Automatics and Robotics,University of Granada), Luis Javier Herrera (Department of Computer Engineering, Automatics and Robotics,University of Granada), Francisco Ortuño (Department of Computer Engineering, Automatics and Robotics,University of Granada) and Ignacio Rojas (Department of Computer Engineering, Automatics and Robotics,University of Granada).

This article presents a study that aimed to identify the stages of bladder cancer based on gene expression data. The dataset used in the study was obtained from the GDC repository and included 406 cases of bladder cancer and 431 files from the TCGA-BLCA project. The study categorized the cases into three classes based on disease stages: Stage 2, Stage 3, and Stage 4. The methodology employed R programming language and the KnowSeq library for the study development. The au- thors identified genes that showed significant differences in expression among the classes and created a matrix of differentially expressed genes (DEG). Machine learning models, including feature selection algorithms and classification models such as KNN and SVM, were constructed to predict the bladder cancer stages. The results revealed that the mRMR feature selection algorithm performed the best, and the 8 most relevant genes were used to build the classification models.

[1100] *Stochastic model of infection with the SARS–COV–2 virus in a small group of individuals indoors*

Igor Derevich (Doctor of Technical Sciences, Prof. of Bauman Moscow State Technical University) and Anastasiia Panova (Postgraduate student of Bauman Moscow State Technical University).

Introduction. COVID-19 virus infection poses a significant danger to the human body due to significant damage to internal organs. The paper presents two models of infection of a small group of individuals, one of which takes into account the random concentration of virions in the atmosphere, and the second – social behavior. The first model considers the initial stage of infection in a small group of individuals, which can be already infected. The spread of infection occurs by airborne droplets as a result of the absorption by the lungs of susceptible members of microdrops, the size of which can be approximately 5 microns, and exhaled by an infected individual. As a result of random movement of individuals in groups, for example, in entertainment venues, supermarkets, waiting rooms, etc., the concentration of virions in the local atmosphere is random. A practical interest of the research lies not only in the modeling of process of infection of susceptible group members as a result of their physical contacts with infected, but also in the prediction of disease development after infected individuals leave the critical zone. The emergence of new viral infections that cause severe consequences in the body of an infected individual leads to the necessity to tighten sanitary and epidemiological standards. However, compliance with these norms in the case of real behavior of a group of individuals is not always fulfilled. This is the most clearly manifested when a group of individuals accidentally moves in places of entertainment, in supermarkets and is disrupted during evacuation from buildings in a panic. The second model of our work takes into account these processes. The norms define a specific critical distance between individuals, with a decrease in which the probability of infection increases. In addition to the critical distance, an important factor is the exposure time of the

susceptible in the radius of active infection near the infected individual. A modern effective tool for assessing the probability of infection in a group of individuals with infected members is the mathematical modeling of the movement of individuals in various indoor conditions, taking into account obstacles. In the literature, much attention is paid to modeling the behavior of large groups of several hundred people during evacuation in panic conditions. The existing mathematical models are constructed by analogy with the mechanics of the movement of hard disks, between which there are repulsive forces when they converge to each other and "friction" forces when in close contact. As a rule, situations with close contact are realized in staged experiments. With a real evacuation from the premises, as can be seen from the photo and video frames, people avoid close contacts and try to maintain social distance even in critical conditions.

Research methods. In the first part of the work, we modified the traditional three–stage model, representing the dynamics of the growth of the concentration of the pathogen SARS–COV-2. The simulation of fluctuations in virion concentration appeared in the local atmosphere and the dynamics of disease development in the body is implemented on the basis of solving a system of stochastic ordinary differential equations (SODE) proposed by modern Runge–Kutta type algorithms. In the atmosphere with fluctuations in the concentration of virions, there is a possibility of intense infection, which will subsequently lead to a severe form of the disease. In our second model, which describes the dynamics of a small group, the social behavior of individuals and their interaction with the room walls are modeled on the basis of an effective potential that increases dramatically when individuals get closer to each other and to borders. The dynamics of social behavior includes a random component of the desired speed of movement, which is modeled by a structured random process.

Conclusion. In the first part of the work, a probability model of infection of individuals with the SARS–COV–2 virus is proposed, the concentration of virions of which in the atmosphere fluctuates near a certain average value. The qualitative difference in the dynamics of infection of body cells in deterministic and random conditions is shown. This model combines the methods of a single description of micro and macro parameters. In the second part of the work, a fundamentally new model of random movement of a group of individuals in a room with obstacles is proposed, taking into account their social behavior in various situations.

The research was carried out at the expense of the grant of the Russian Science Foundation No. 23-29-00243, https://rscf.ru/project/23-29-00243/

[1151] *The Coherent Multi-Representation Problem with Applications in Structural Biology*

Antonio Mucherino (IRISA, University of Rennes).

We introduce the Coherent Multi-Representation Problem (CMP), whose solutions allow us to observe simultaneously different geometrical representations for the vertices of a given simple graph. The idea of graph multi-representation extends the common concept of graph embedding, where every vertex can be embedded in one unique domain that is common for each vertex. In the CMP, the same vertex can instead be represented in multiple ways, and the main aim is to find a general multi-representation where all the involved variables are “coherent”. We prove that the CMP extends a geometrical problem known in the literature as the distance geometry problem, and we show some very preliminary computational experiments performed with a new Java implementation specifically conceived for graph multi-representations.

[1236] *Comparison of VCFs generated from different software in the evaluation of variants in genes responsible for rare thrombophilic conditions*

Radek Vrtel (Institut of Medical Genetics, University Hospital and Palacky University Olomouc), Petr Vrtel (Institut of Medical Genetics, University Hospital and Palacky University Olomouc) and Radek Vodicka (University Hospital and Palacky University).

Introduction. Second-generation sequencing techniques, so-called massively parallel sequencing (MPS), are rapidly gaining ground in medical genetics for the diagnosis of genetic diseases due to their speed and capacity. With the ability to simultaneously screen multiple patients for multiple genetic targets, they are replacing previously used molecular diagnostic methods, which was mainly Sanger sequencing. As part of the implementation and validation of an optimal diagnostic approach based on high-throughput sequencing in the diagnosis of the rare thrombophilic conditions of protein S (PS), protein C (PC) and antithrombin (AT) deficiency, we compared data from three different software tools - Torrent Suite, Ion Reporter and NextGene - to compare their performance and accuracy in the analysis of each sequence variant detected.

Methods. The Ion Torrent S5 and Ion Chef equipment was used for amplicon library preparation and targeted sequencing. In order to compare the quality of the three different evaluation software, 10 patients were selected from one sequencing run in which patients were indicated for PS, PC and AT deficiency, but also for RASopathy (all signed a consent to use the results anonymously). Patients were anonymized and only variants detected in the PROS1, PROC and SERPINC1 genes were evaluated.

The study compared variants in terms of their presence/absence in a given VCF cohort. The VCFs compared were those generated from Torrent Suite software using the integrated variantCaller module; VCFs generated from BAM files in Ion Reporter software and VCFs generated using NextGene software. The settings of variantCaller in Torrent Suite and Ion Reporter are left default.

Results. All software successfully detected the causal variant SERPINC1:c.79T>C (p.Trp27Arg). In these patients, the three software detected 16, 19, and 27 variants in the PROS1 gene; 17, 17, 19 variants in the PROC gene; and 15, 15, 16 variants for the SERPINC1 gene in their baseline settings (shown in Table). Further detailed data will be presented in a poster.

Results of individual software: The NextGene software was the only one to detect INDEL variants present in homopolymeric regions: in the PROS1 gene c.1870+84delA (in patients 3, 4, 6, 8), c.602-23delT (in patients 4, 5, 6, 8, 10), c.469+25delC (in patients 4 and 8), c.1156-9dupT (in patient 10). In the PROC gene, c.1212delG (p.Met406TrpfsTer15) (in patients 3 and 9). In the SERPINC1 gene, this software detected the variant c.409-13_409-12delTT in patient 9. The frequency of alternative alleles for these variants ranged from 20.6-36.9%. The presence of INDEL variants in homopolymeric regions in multiple patients in the same sequencing run with relatively low frequency suggests that these are sequencing artefacts.

The analysis in Ion Reporter software differed from the other compared software by detection of variants in PROS1 gene c.347-47delT in patients 1 and 7 (frequency of alternative allele 50.98 and 55.49%) and c.939A>T (p.Leu313Phe) in patients 3 and 5 (frequency of alternative allele 20.91 and 21.97%). According to the Integrative Genomics Viewer IGV, the variant was present in all patients within a given sequencing run, but it is not listed in the VarSome genomic viewer. This INDEL located in the A/T rich region has low quality, which is the lowest ranked for VC by the Ion Reporter, so it is probably a sequencing artefact. Using the Torrent Suite software, a variant in the PROS1 gene c.1156-9_1156 was detected in patient 9 with an alternative allele frequency of 43.96% compared to the other two software. The same variant (c.1156-9dupT) was detected in patient 10 by NextGene software with an alternative allele frequency of 34.7%. This is an INDEL in the A/T rich region - according to IGV present in all patients in a given sequencing run, not listed in VarSome - it is most likely a sequencing artefact.

Discussion For INDEL variants, especially in homopolymeric regions, we observed the highest number of missdetected variants in NextGene software. For data generated from the Ion Torrent platform, software from the same provider seems to be more suitable, mainly because of the quality of the false positive filtering. For further evaluation of the validity of the software used, it will be necessary to expand the cohort of patients examined.

Tab. The number of variants detected by the selected software in each patient.

[1267] *Developing a workflow to gain insights from scRNA-seq data for drug development : a case-study with human pancreatic adenocarcinoma*

Octavio Morante-Palacios (Clarivate Consulting Services), Ester Gil Vazquez (Clarivate Consulting Services), Oleg Deryagin (Clarivate Consulting Services) and Cecilia Klein (Clarivate Consulting Services).

Single-cell RNA-seq (scRNA-seq) analysis is a powerful tool for gaining insights into cancer biology and drug development. In this study, we obtained a public scRNA-seq dataset of pancreatic ductal adenocarcinoma (PDAC) patients with different chemotherapy combinations. We further investigated the expression changes produced in tumor cells and in the tumor microenvironment using Clarivate mechanism of action reconstruction workflow based on network algorithms (overconnectivity and causal reasoning) and Clarivate CBDD (Computational Biology for Drug Discovery) tools. Our analysis revealed potential regulators involved in these expression changes and their associated perturbed pathways by cell type. We also assessed differences in gene expression between responders and non-responders to understand potential regulators mediating drug resistance. Our results provide a collection of differentially expressed genes and perturbed pathways for treated PDAC patients compared to non-treated patients, as well as responders compared to non-responders. We contextualized our findings with PDAC-related biomarkers and targets from Clarivate databases, such as MetaBase™ and Cortellis Drug Discovery Intelligence. Our approach can offer valuable insights for preclinical and clinical studies, such as patient stratification, understanding cell-cell interactions, assessing heterogeneity, and investigating mechanism of action. Overall, our study highlights the potential of scRNA-seq analysis combined with

network algorithms, Clarivate CBDD and manually curated proprietary databases to better understand pancreatic adenocarcinoma, which could ultimately inform the development of more effective treatment for this disease.

[1279] *Predictive model for Age-Related Macular Degeneration Response to Anti-VEGF treatment*

Álvaro Pérez Sala (Fundación Riojasalud), Rafael Pelaez (Center for Biomedical Research of La Rioja (CIBIR)), Ana Isabel Oca Lázaro (San Pedro Hospital), Ángela Villanueva Martinez (Fundación Rioja Salud) and Ignacio M. Larráyoz Roldán (Fundación Riojasalud).

During the last years, signature-based predictive models have been applied to different pathologies obtaining promising results. Those models take special consideration in situations where there is not clear clinical guidelines to choose between different treatments in a specific case for early application. Moreover, predictive models focused on treatment response play an indispensa-ble role on personalized medicine. Age-related macular degeneration (AMD) is an incurable disease associated with aging that destroys sharp and central vision. Most widespread treatment is intravitreal injection of anti-vascular endothelial grow factor (VEGF) agent to prevent choroidal neovascularization in those pa-tients. However, it is not effective on absolutely all patients since many of them do not show significant response to the treatment after three doses. The develop-ment of a predictive treatment response model to AMD would be very helpful to decide which treatment is suitable for each patient. Here, we present a novel clas-sification model based on a signature composed of 4 mRNAs and 1 miRNA, ob-tained from PBMCs, that predicts the response to ranibizumab with high accu-racy (Area Under the Curve of the Receiver Operating Characteristic curve = 0.968), before treatment. Our classification model represents a robust screening method to identify those patients with poor response to anti VEGF. Furthermore, this model in combination with other information, such as specific baselines char-acteristics, could help to establish a treatment plan on the first visit.

[1590] *Insulin sensitivity and Insulin Resistance and Diabetes prediction by applying artificial intelligence techniques on genes obtained from expression analysis differential*

Jesús María González-Martín (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín), Francisco Rodríguez-Esparragón (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín), Bernardino Clavo (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín), Laura Beatriz Torres-Mata (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín), Sara Estefania Cazorla-Rivero (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín) and Estrella Gómez-Bentolila (Research Unit, Hospital Universitario Gran Canaria Doctor Negrín).

Currently, 463 million adults have diabetes and 374 million have impaired glucose tolerance. Insulin is a powerful pleiotropic hormone that affects processes such as cell growth, energy expenditure, and carbohydrate, lipid, and protein metabolism. On the other hand, skeletal muscle is the main site for insu-lin-dependent glucose excretion. The molecular mechanisms by which insulin regulates muscle metabolism and the underlying defects that cause insulin re-sistance have not been fully elucidated. The objective of this study is to perform an analysis of microarray data to find differentially expressed genes. The analysis has been based on the data of a study deposited in Gene Expression Omnibus (GEO) with identifier "GSE22309" and whose title is "Human skeletal muscle expression data". The selected data contains samples from three types of patients after taking insulin treatment: patients with diabetes (DB), patients with insulin sensitivity (IS), and patients with insulin resistance (IR). Once the 20 genes expressed differentially between the three possible comparisons were obtained (DB vs IS, DB vs IR and IS vs IR), this data set has been used to develop predictive models through Machine Learning techniques to classify patients with respect to the three categories mentioned previously. All the techniques used present an accuracy superior to 80%, reaching almost 90% when unifying the categories IR and DB.

[1742] *Validation of Height-for-Age and BMI-for-Age Z-scores Assessment using Android-based Mobile Apps*

Valerii Erkudov (St. Petersburg State Pediatric Medical University), Sergey Lytaev (St. Petersburg State Pediatric Medical University), Kenjabek Rozumbetov (Berdakh Karakalpak State University), Andrey Pugovkin (St. Petersburg State Pediatric Medical University), Azat Matchanov (Berdakh Karakalpak State University) and Sergey Rogozin (St. Petersburg State Pediatric Medical University).

The WHO's standards for analysis and presentation of anthropometric data is widely recognized as the best system. This study was aimed to assess the agreement of calculation of height-for-age and BMI-for-age Z-scores according to WHO growth charts using of two randomly selected mobile apps in two cohorts of European and Asian subjects. In 1,347 adolescents aged 13 to 17 years, boys and girls, living in St. Petersburg, Northwest of the Russia (Europe-an cohort, 663 subjects) and Nukus, Uzbekistan (Asian cohort 684 subjects) measured body weight and stature. Each child's height-for-age Z-score and BMI-for-age Z-scores were calculated based on WHO Child Growth Standards PC version WHO AntroPlus software and mobile applications. It was performed a Blend-Altman analysis to assess the consistency of Z-scores calculated using WHO AntroPlus software and two mobile apps. The results of the evaluation of the consistency of height-for-age and BMI-for-age Z-scores obtained using WHO AntroPlus software and mobile applications showed that Android-based mobile applications systematically overestimate Z-scores. However, this error is much smaller than any clinically significant deviations of the physical development evaluation parameters both in the European and Asian cohort volunteers. Thus, it can be claimed that it is possible to use Android-based mobile applications for growth monitoring and BMI and WHO AntroPlus software for children from 5 to 19. This is especially relevant to the need for «field» monitoring of children's growth and development performed by general practitioners visiting patients at home.

[1758] *Predicting and detecting coronary heart disease in patients using machine learning method*

Michal Wos (Department of Medical Informatics and Statistics with e-learning Lab. Medical university of lublin), Bartłomiej Kiczek (Department of physics University Curie-Skłodowska University of lublin) and Bartłomiej Drop (Department of Medical Informatics and Statistics with e-learning Lab. Medical university of lublin).

Machine learning creates new opportunities for medicine and public health, especially in the field of helping medical workers. The prospect of generating hints for the diagnosis of a particular disease for doctors is widely considered in the field of machine learning. The work will only present the possibility of predicting coronary heart disease, which is classified as a civilization disease, threatens an increasing number of people in the world. According to the data of the Institute for Health Metrics and Evaluation (IHME, 2020), in 2017, 3.5% of people in the European Union suffered from coronary heart disease (I20-I25, Z82.4-Z82.49 according to ICD-10). The IHME research also shows that the percentage of patients is increasing year by year in the European population. Due to the growing number of patients, additional problems such as: staff shortages and increasing queues to specialist doctors can now be observed. The solution to these issues is found in domain information systems that use machine learning to detect or predict the possibility of contracting the disease. A database of 11,000 patients was used for the study. A preliminary analysis of the onset of Coronary Heart Disease was carried out. The main factors were age, LDL cholesterol level, systolic and diastolic blood pressure. Three of these factors (which can be eliminated through exercise and a proper diet) account for the highest percentage of the probability of developing Coronary heart disease. In total each record of the dataset contained 13 features (which were analyzed by the model) such as BMI, weight, total cholesterol, years of smoking, however the analysis of these factors did not significantly contribute to the prediction of ischemic heart disease. In the search of the best fit we browsed several families of regressors. Starting from different variants of the linear model, with various regularisation factors, to support vector machines, ensemble learning (RandomForest) and tree-based solutions with gradient boosting. It turned out that extreme gradient boosted trees (XGBoost) proved to provide best results. For finding the optimal solution a grid search was performed, where parameters like number of estimators, depth, shrinkage parameter, alpha (L1 regularization term), lambda (L2 regularization term), subsampling rate and column sampling (internal feature selection of the model) were optimized. The final model was trained using 5-fold cross validation, with mean Pearson’s R2 exceeding 0.97 and root mean squared error equal 0.013. The model was also validated on a separate test set and presented similarly good results. After the model was trained, we performed a series of backward searches, such as a visual analysis of the tree models, model estimated weights and an additional LIME-based (Local Interpretable Model-agnostic Explanations) explanation. Additionally, such tree models provide a straightforward method of making decisions, which is also very informative and educative from human perspective. It has been found that merely three features are good-enough to precisely estimate the likelihood of coronary heart disease incidence. The remaining features provided minor, in some cases even negligible, contribution to the total likelihood. The obtained results look promising in terms of a deployment of this particular model, but also similar models aimed in different diseases, in software supporting public health professionals. Such machine learning based solution can serve as an as an assistant for a physician, but also can generate health warnings to patients during various examinations, where their data is gathered.

[1799] *Three-dimensional representation and visualization of high-grade and low-grade glioma by Nakagami imaging*

Orcan Alpar (University of Hradec Kralove) and Ondrej Krejcar (University of Hradec Kralove).

Three-dimensional (3D) visualization of the brain tumors reconstructed from the two-dimensional (2D) magnetic resonance imaging (MRI) sequences plays an important role in volumetric calculations. The reconstructions are usually executed using the fluid attenuated inversion recovery (FLAIR) sequences, where the whole tumors appear brighter than the healthy surrounding tissues. Without any processing; however, reconstruction results might be inconclusive; therefore, we propose a mathematical m-parametric Nakagami imaging for highlighting the lesions. The raw 2D FLAIR MRI images are taken from BraTS 2012 dataset and the highlighted images are generated by the Nakagami imaging. The information on the MRI slices is compiled in three-layered Nakagami images for better visualization of the high-grade and low-grade glioma in 3D space. By the flexible m-parametric design, on the other hand, the reconstructed images might easily be adjusted according to the GT images for precise representation.

[1853] *Revealing the RBP regulome in hepatocellular carcinoma via consensus GRN inference*

Mateusz Garbulowski (Department of Biochemistry and Biophysics, Stockholm University, Science for Life Laboratory, Solna, Sweden), Riccardo Mosca (Department of Microbiology, Tumor, and Cell Biology, Karolinska Institute, Science for Life Laboratory, Solna, Sweden), Carlos J. Gallardo-Dodd (Department of Microbiology, Tumor, and Cell Biology, Karolinska Institute, Science for Life Laboratory, Solna, Sweden), Claudia Kutter (Department of Microbiology, Tumor, and Cell Biology, Karolinska Institute, Science for Life Laboratory, Solna, Sweden) and Erik L. L. Sonnhammer (Department of Biochemistry and Biophysics, Stockholm University, Science for Life Laboratory, Solna, Sweden).

RNA binding proteins (RBPs) are proteins that play a critical role in controlling various mechanisms of gene expression by binding hundreds of RNA targets, often via a known RNA binding domain (RBD) [1]. They are involved in a wide repertoire of post-transcriptional processes in human cells (Figure 1). Among others, RBPs are associated with mRNA decay, splicing, stabilization, translation and transport which makes them a highly influential group of proteins. In cancer cells, RBPs are involved in various mechanisms of carcinogenesis and drug resistance [2]. Recent research by Cen et al. highlighted the role of RBPs in breast cancer drug resistance [3]. Yet another study has described the pro-tumorigenic role of a cytoplasmic RBP, hnRNPK [4]. Thus, there is an urgent need to investigate a poorly explored regulome of RBPs in cancer. In this study, we focused on the analysis of hepatocellular carcinoma (HCC) that still remains a challenging therapeutic task.Over the past years, gene regulatory networks (GRNs) have been successfully applied to reveal the cancer regulome. The main aim of GRN is to characterize interactions between regulators and their targets in biological systems. However, no gold standard method was established to tackle this dilemma in an accurate way. Another issue is that biological data are often noisy that causes a high risk of capturing false positive interactions and introduces heterogeneity in the analysis. Importantly, it has been evaluated that the performance of the same GRN inference methods vary between diverse datasets [5]. Furthermore, a work by Seçilmiş et al. [6] has shown that knowledge about a perturbation profile is essential to infer a meaningful GRN. Thus, experiments that allow for controlled perturbation of gene expression such as shRNA knockdown followed by RNA-seq (shRNA-seq) shall be extensively used for GRN inference. Hence, in this work, we applied a consensus approach to infer a GRN from shRNA-seq data by combining popular and effective regression and machine learning methods. To evaluate the performance of our methodology, we generated synthetic data based on the properties of shRNA-seq ENCODE data sets. Moreover, we performed a comprehensive validation for identified RBP-RBP interactions. To reveal the RBP regulome for a liver cancer, we used publicly available shRNA-seq of HCC cell line (HepG2) stored by the ENCODE project [7]. In addition, we incorporated shRNA-seq of chronic myelogenous leukemia cell line (K562) that was used to generate a control network. For both cell lines, 232 RBPs were investigated. To infer a consensus network, we used the GeneSPIDER framework [8] from which we employed 12 methods: least square, normalized least squares, ridge regression with least square cutoff, lasso, elastic net, logistic regression, support vector machine, z-score, decision trees, bagging trees, logit boost trees and gentle boost trees. Complementary methods were combined, such as lasso and elastic net, resulting in eight GRN networks used for the consensus approach. The final consensus network for HepG2 contained 892 links which were inferred by at least two methods. To evaluate the consensus approach, we performed a benchmarking of synthetically generated ENCODE-like data where the gold standard network was pre-declared. Our results showed that a consensus approach reached significantly higher precision in comparison to using a single GRN inference method (Student's t-test $P < 0.05$). Furthermore, the consensus approach removed much more false positives than true positives from the synthetic GRNs. Next, in order to validate the HepG2 GRN, we employed information from diverse databases and resources, referred to as validation features, such as 1) K562 cell line GRN estimated in the same way as HepG2, 2) GTEx with healthy gene expression from liver tissue, 3) TCGA with liver HCC (LIHC) gene expression from patients and impact on their survival, 4) EURBPDB with number of RBPs reported in cancer-relevant literatures and differentially expressed in cancers comparing to healthy tissue, 5) FunCoup5 with functional associations determining physical or regulatory interactions, 6) Kutter et al. supplementary material with a list of canonical and non-canonical RBPs and number of literature mentions about a gene being RBP and RAP-seq experiment

results for selected RBPs [9, 10], and 7) ENCODE with eCLIP-seq data that represent a landscape of RBPs-RNA interactions. For FunCoup5, absent association scores were estimated using decision trees learning based on the training (cross validation accuracy 77%) with all the validation features where interactions contained the FunCoup5 score. Lastly, an experimental validation will be performed on selected RBPs. Based on the HepG2 GRN, we identified an RBP regulome (Figure 2) that is enriched with terms related to liver carcinogenesis using the DisGeNET resource, which contains a large collection of gene-disease associations. Preliminary results demonstrate that the obtained GRN reveals known and novel interactions for HCC (Figure 2). For instance, one of RBP-RBP associations is IGF2BP1-PTBP1, confirmed with eCLIP-seq and RAP-seq [10], where these genes are more co-expressed in LIHC than GTEx control. This novel interaction was not present in the FunCoup5 database but follows the pattern of highly functional associations estimated with decision tree learning. Another example can be EWSR1-PES1 having a link in three GRNs and high FunCoup5 score. Interestingly, this interaction significantly ($P < 0.05$) affects survival of LIHC patients but was not frequently mentioned in the literature as differentially expressed in cancers. Furthermore, we have validated a list of hubs, i.e., genes with high connectivity, for a full consensus GRN by intersecting their targets with eCLIP-seq ENCODE data. We observed that AQR, SUB1 and U2AF2 exhibit a significant overlap (hypergeometric test P value < 0.05) with the eCLIP-seq peaks and share over 75% of their targets. Taken together, we believe that the RBP regulome of HCC may serve as a resource of therapeutic hypotheses that can be used in the future to improve treatment.

References

[1] Gebauer, F., Schwarzl, T., Valcárcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. Nature Reviews Genetics, 22(3), 185-198. [2] Pereira, B., Billaud, M., & Almeida, R. (2017). RNA-binding proteins in cancer: old players and new actors. Trends in cancer, 3(7), 506-528. [3] Cen, Y., Chen, L., Liu, Z., Lin, Q., Fang, X., Yao, H., & Gong, C. (2023). Novel roles of RNA-binding proteins in drug resistance of breast cancer: from molecular biology to targeting therapeutics. Cell Death Discovery, 9(1), 52. [4] Mucha, B., Qie, S., Bajpai, S., Tarallo, V., Diehl, J. N., Tedeschi, F., ... & Diehl, J. A. (2022). Tumor suppressor mediated ubiquitylation of hnRNPK is a barrier to oncogenic translation. Nature Communications, 13(1), 6614. [5] Shen, B., Coruzzi, G., & Shasha, D. (2023). EnsInfer: a simple ensemble approach to network inference outperforms any single method. BMC bioinformatics, 24(1), 114. [6] Seçilmiş, D., Hillerton, T., Tjärnberg, A., Nelander, S., Nordling, T. E., & Sonnhammer, E. L. (2022). Knowledge of the perturbation design is essential for accurate gene regulatory network inference. Scientific reports, 12(1), 16531. [7] Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., ... & Cherry, J. M. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic acids research, 48(D1), D882-D889. [8] Tjärnberg, A., Morgan, D. C., Studham, M., Nordling, T. E., & Sonnhammer, E. L. (2017). GeneSPIDER–gene regulatory network inference benchmarking with controlled network and data properties. Molecular BioSystems, 13(7), 1304-1312. [9] Søndergaard, J. N., Sommerauer, C., Atanasoai, I., Hinte, L. C., Geng, K., Guiducci, G., ... & Kutter, C. (2022). CCT3-LINC00326 axis regulates hepatocarcinogenic lipid metabolism. Gut, 71(10), 2081-2092. [10] Atanasoai, I., Papavasileiou, S., Preiß, N., & Kutter, C. (2021). Large-scale identification of RBP-RNA interactions by RAPseq refines essentials of post-transcriptional gene regulation. bioRxiv, 2021-11.

Fig. 1: Post-transcriptional roles of RBPs. Fig. 2: Overview of top scoring links from the consensus GRN. A. Validation features and their scores for top interactions selected based on DisGeNET enrichment towards liver cancer-related terms. P values are marked as ns ($P > 0.1$), * ($P \leq 0.1$), ** ($P \leq 0.05$), *** ($P \leq 0.01$) and **** ($P \leq 0.001$). For co-expression healthy (H) and cancer (C) tissues are marked. For intersection with the control K562 GRN, common (c) and uncommon (unc) links are marked. For RBD canonical (c) and non-canonical (nc) domains are marked. Lack of information is marked as Ø or empty tile. B. GRN for top interactions with categorized low and high FunCoup5 score, and its estimated low* and high* category.

[1884] *Conserved Water Networks Identification in Proteins Using Density Clustering Approaches on Positional and Orientational Data*

Urban Bren (Faculty of Chemistry and Chemical Technology, University of Maribor), Marko Jukic (Faculty of Chemistry and Chemical Technology, University of Maribor), Jelena Tosovic (Faculty of Chemistry and Chemical Technology, University of Maribor) and Domagoj Fijan (Department of Chemical Engineering, University of Michigan).

This work describes the development and testing of a method for the identification and classification of conserved water molecules and their networks from molecular dynamics (MD) simulations. The conserved waters in the active sites of proteins influence protein−ligand binding. Recently, several groups have argued that a water network formed from conserved waters can be used to interpret the thermodynamic signature of the binding site. We implemented a novel methodology to categorize water molecules extracted from the MD simulation trajectories using clustering approaches. The main advantage of our methodology as compared to the current state-of-the-art approaches is the inclusion of information on the orientation of hydrogen atoms to further inform the clustering algorithm and to classify the conserved waters into different subtypes depending on how strongly certain orientations are preferred. This information is vital for

assessing the stability of water networks. The newly developed approach is described in detail as well as validated against known results from the scientific literature including comparisons with the experimental data on thermolysin, thrombin, and Haemophilus influenzae virulence protein SiaP as well as with the previous computational results on thermolysin. We observed excellent agreement with the literature and were also able to provide additional insights into the orientations of the conserved water molecules, highlighting the key interactions which stabilize them. The source code of our approach, as well as the utility tools used for visualization, are freely available on GitHub.

[1997] *Motion control of a robotic lumbar spine model*

Thuanne Paixão (University of Acre), Ana Beatriz Alvarez (University of Acre), Ruben Florez (University of San Antonio Abad del Cusco) and Facundo Palomino-Quispe (University of San Antonio Abad del Cusco).

The study of the movement of the lumbar spine vertebrae is classified as a relevant topic for research considering the possibility of exploring pathological dysfunctions of the spine. In this paper, we present the development of a motion control for a lumbar spine model. The model is being represented by a 2 DOF (Degrees of Freedom) manipulator robot, which represents the motion of two lumbar vertebrae. For the computer simulations, the mathematical model is being used. The dynamic mathematical model of the manipulator is being considered, based on the Lagrange approach. Preliminary simulation results show that the implemented conventional controller robustly follows the given references. The signal references represent the angles of the lumbar spine vertebrae, thus guaranteeing the planned movement.

[2083] *Targeted Next Generation Sequencing of a custom capture panel to target sequence 112 cancer related genes in breast cancer tumours ERBB2 positive from Lleida (Spain)*

Ana Velasco (HOSPITAL ARNAU DE VILANOVA DE LLEIDA), Izaskun Urdanibia (IRB Lleida) and Serafín Morales (Hospital Universitari Arnau de Vilanova de Lleida).

The diagnosis, prognosis and treatment of breast cancer is based on clinical examination in combination with imaging and confirmed by pathology determination of histology, tumour grade, oestrogen receptor (ER), progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2/ERBB2) and proliferation marker Ki67 in core biopsies (Cardoso, Kyriakides et al. 2019). These biological markers are evaluated by immunohistochemical methods (IHC) using a standardised assessment methodology, e.g. Allred score or H-score, thereby HER2 is defined as positive by IHC (3+) when more than 10% of the cells harbour a complete membrane staining (protein is overexpressed), and by in situ hybridization (ISH) if the number of HER2 gene copies is >6. HER2 is part of the epidermal growth factor (EGF) family that stimulates cell growth and differentiation. Between 15-30% of invasive breast cancers have HER2 gene amplifications resulting in HER2 protein overexpression, which leads to an increase in tumor growth and aggressiveness (Wang 2017). Anti-HER2 drug Trastuzumab is a humanized monoclonal antibody targeting HER2 receptor, which was approved for use in 1998. Trastuzumab combined with chemotherapy (ChT) in patients with HER2 overexpression/amplification approximately halves the recurrence and mortality risk, compared with ChT alone (Ross, Slodkowska et al. 2009). The mechanisms of action of Trastuzumab have not been clearly defined, but likely include extracellular mechanisms involving antibodydependent cellular cytotoxicity (ADCC), and intracellular mechanisms involving apoptosis and cell cycle arrest as well as inhibiting angiogenesis, and preventing DNA repair following chemotherapy-induced damage. HER2 positive patients in Lleida are diagnosed and treated in Hospital Universitari Arnau de Vilanova (Lleida) according guidelines mentioned above. Even though such homogeneous group, every patient has their own prognosis based on different features, some of which genetic involved. With that aim, we have implemented a custom NGS panel comprising three probe subgroups for testing targeted mutations, copy number alterations and translocations in tumors with hormonal receptor (HR) and HER2 status previously assessed by immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH).

[2118] *GeneCaRNA: A new world of non-coding RNAs for disease decipherment*

Doron Lancet (Weizmann Institute of Science).

Background: Most clinical geneticists merely explore the ~2% of the human genome that harbors the ~20,000 protein-coding genes for therapeutic and diagnostic candidate searches. Even if researchers want to decipher disease using whole genome sequencing for non-coding variant search, the gene-centric ncRNA world seems to lag the protein-coding equivalent, e.g., only ~8,000 ncRNA genes with an HGNC-approved symbol (PMID:32090359). Over time, transforming this information into a comprehensive set of genome-mapped ncRNA genes became crucial. Hence, we developed

GeneCaRNA (PMID:33676929) under the GeneCards Suite. Methods: For developing GeneCaRNA, we clustered overlapping transcript entries from all sources (17 transcript sources in RNAcentral plus HGNC, NCBI Gene, and Ensembl), resulting in complete coverage of genomic gene positions using genomic coordinates. GeneCaRNA genes are cataloged into 17 categories based on similar partitions in 4 major data sources. Results: Until GeneCaRNA was developed, only ~32,000 ncRNA genes existed in primary gene sources. We defined additional ~188,000 records as novel Transcripts Inferred GeneCaRNA genes (TRIGGs). Each GeneCaRNA gene is shown on a web 'card,' including affiliated transcripts, category affiliation, and affiliated diseases from MalaCards (PMID:28830434). The most prominent type is piRNA (109,820 genes), amounting to ~50% of all GeneCaRNA genes; the second prominent type is lncRNA, with 75,839 genes, ~35% of the compendium; the rest encompass the remaining 15 types (sRNA, vaultRNA, guideRNA, catalyticRNA, scaRNA, scRNA, yRNA, rRNA, tRNA, snoRNA, snRNA, miRNA, modelRNA, miscRNA, and srpRNA). Thus, the GeneCards Suite now benefits from comprehensive ncRNA data access, including VarElect (PMID:27357693), our variant-disease decipherment tool, which now addresses whole genome sequencing (WGS) results. Conclusion: The strength of GeneCaRNA is mapping all ncRNA transcripts to genome coordinates, making WGS usable for linking DNA variants to diseases. GeneCaRNA is becoming a popular gateway for WGS disease interpretation, as attested by inclusion in the UCSC browser and in others deciphering unsolved diseases, e.g., a Gastric Cancer (PMID:35571069) and an Alzheimer's disease instance (PMID:35147545). It also supported the creation of an improved piRNA database, including relationships to human diseases (PMID:35667080). Thus, we firmly believe that GeneCaRNA will greatly facilitate exploring the non-coding RNA world to decipher unsolved diseases.

[2158] *Competitive Analysis of 5 S AND 16 S bacterial RNA in the Phylogeny of bacteria*

Michael Sadovsky (Institute of computational modelling of SB RAS) and Yulia Ovchinnikova (Siberian Feceral University).

Statistically revealed inner structuredness of bacterial vs. chloroplast phototsynthetic genes is studied. To do it, we analyzed the cyanobacterial genes responsible for a light consumption. A sounding difference in the spatial pattern specific for chloroplast photosystem genes, and bacterial photosystem genes is found. Thus, the bacterial genes yield another type of symmetry of the distribution of the genes converted into triplet frequency dictionaries in 63-dimensional Euclidean space of triplets.

[2271] *An Algorithm for Pairwise DNA Sequences Alignment*

Veska Gancheva (Technical University of Sofia) and Hristo Stoev (Technical University of Sofia).

A challenge in data analysis in bioinformatics is to offer integrated and modern access to the progressively increasing volume of data, as well as efficient algorithms for their processing. Considering the vast databases of biological data available, it is extremely important to develop efficient methods for processing biological data. A new algorithm for arranging DNA sequences based on the suggested CAT method is proposed, consisting of an algorithm for calculating a CAT profile against the selected reference sequences and an algorithm for com-paring two sequences, based on the calculated CAT profiles. Implementation steps, inputs and outputs are defined. A software implementation of the proposed method for arranging biological sequences CAT has been designed and developed. Experiments have been carried out using different data sets to align DNA sequences based on CAT method. An analysis of the experimental results have been done in terms of collisions, speed and effectiveness of the proposed solutions.

[2323] *Speeding up simulations for radiotherapy research by means of machine learning*

Luis Javier Herrera (University of Granada), Ignacio Rojas (University of Granada), Francisco Carrillo Pérez (Universidad de Granada), Alberto Guillen (University of Granada), Isabel Fernández (University of Granada) and Carmen Ovejero (Kerma S.L.).

Radiotherapy is one of the most widely used treatments for cancer by irradiating the tumor volume. However, one of its disadvantages is that healthy tissue is also affected, producing various side effects. For this reason, studies are carried out beforehand to determine the dose to be administered in each case, to avoid this damage and to ensure that the dose received by the tumor is the correct one. These studies are carried out both using simulations and with routine machinery procedures using a mannequin that simulates the area to be treated. In this work a way of speeding up the previous study process is tackled, starting from simulated data whose optimized obtaining will be the objective of this work. The PENELOPE Monte Carlo simulation code is used to recreate the process and obtain the necessary previous

data. Subsequently, regression models are applied to obtain the values of interest and accelerate the procedure, reducing, in addition, the energy consumption and storage required while obtaining very accurate approximations.

[2379] *HEALTH CARE AND DISEASES*

Aditi Katake (New Vision University , Georgia).

PURPOSE - Health for all is an ideal goal that all governments aspire to reach. The purpose of this paper is toassess the definitions of the key terms used to better appreciate the role of the WHO member states in their efforts to achieve improved healthcare systems that suit each nation's particular needs.

[2503] *Role of parallel processing in brain magnetic resonance imaging*

Ayca Kirimtat (University of Hradec Kralove) and Ondrej Krejcar (University of Hradec Kralove).

Parallel processing is a procedure for making computation of more than a pro-cessor to overcome the difficulty of separate parts of an overall task. It is really crucial for some medicine-related tasks since the method provide time-efficient computation by a program, thus several calculations could be made simultane-ously. Whereas, magnetic resonance imaging (MRI) is one of the medical imag-ing methods to show form of an anatomy and biological progressions of a hu-man body. Parallel processing methods could be useful for being implemented in MRI with the aim of getting real-time, interventional and time-efficient ac-quisition of images. Given the need of faster computation on brain MRI to get early and real-time feedbacks in medicine, this paper presents a systematic re-view of the literature related to brain MRIs focusing on the emerging applica-tions of parallel processing methods for the analysis of brain MRIs. We investi-gate the articles consisting of these kernels with literature matrices including their, materials, methods, journal types between 2013 and 2023. We distill the most prominent key concepts of parallel processing methods.

[2555] *An efficient algorithm for detecting mutually exclusive patterns across multiple sets of genomic mutations*

Siyu He (Xi'an Jiaotong University), Jiayin Wang (Xi'an Jiaotong University), Zhongmeng Zhao (Xi'an Jiaotong University) and Xuanping Zhang (Xi'an Jiaotong University).

Mutual exclusive patterns of somatic mutations are reported as important biomarkers in cancer genomics and suggested valuable in guiding cancer treatment. However, detecting mutual exclusive patterns form mutation data is an NP-hard problem. On large-scale dataset, the existing approaches either limit on pair-wise patterns or largely rely on priori knowledge. In addition, the existing approaches often consider genotype data, which loss the information on tumor clonality. In this paper, we present an efficient algorithm for detecting mutual exclusive patterns on multiple mutation sets. Different from the existing approaches, the proposed algorithm focuses on both the similarity within a subset of mutations and the mutual exclusive patterns among the subsets. A degree measures mutual exclusion is defined as optimization objective. The subsets of mutations are considered to present a high-degree of similarity within a subset, while a significant mutual exclusion to other subsets. To facilitate the computation, the proposed algorithm is implemented on a fuzzy strategy, which enables to consider multiple clustering centers simultaneously for one mutation. We conducted a series of experiments to verify the performance of the algorithm on simulation datasets and TCGA mutation calls, while compared to MEGSA. According to the results, the algorithm outperformed MEGSA under main configurations and presented acceptable performance on others. We identified some mutual exclusive patterns on TCGA datasets, which were also supported by literatures.

[2557] *Identification of Novel Anti-inflammatory Agents by Molecular Docking Studies*

Sadaf Naeem (University of Karachi), Atia Shaheen (University of Karachi) and Sadaf Naeem (University of Karachi).

Computational methods that predict the structure and specificity of ligand-protein interactions can yield deep insight into the structural biology of many biological pathways. Molecular docking is computational tool commonly applied in drug discovery projects and fundamental biological studies of protein-ligand interactions. Traditionally, molecular docking is used to screen a collection of small molecules against a receptor and predict the binding affinity of the ligand-receptor complex and thus identify potential active ligands. Here, chemical constituents from selected Indonesian medicinal plants have been virtually screened for their efficacy by inhibiting COX-2 - an enzyme play important roles in pain, inflammation

and cancer. Molecular docking studies were performed using Molegro Virtual Docker (MVD) software on 202 compounds found in ten selected Indonesian traditional medicinal plants (Ceibapentandra (L.)Gaertn, Acorus Calamus L, Ipomoea Batatas (L.), Ananas Comosus (L.) Merr, Citrus Aurantifolia, Lantana Camara, Morinda Citrifolia L, Moringao Leifera Lamk, Plantago Major L, Rosa Damascena Miller) against COX-2 enzyme. These docking studies revealed high binding affinity of compounds - acubin, apigenin, syringin and salicylic acid against cyclooxygenase-2, (PDB ID 4COX), thus, could be potent inhibitors of COX-2 and can be further investigated for in-vitro and in-vivo activity. These molecular docking calculations not only explore the probable binding conformations of compounds within the active site of protein but it provide further useful information in understanding the structural & chemical features of COX-2 inhibitors in designing and finding new potential inhibitors.

[2666] *Hybradization of Emperical Mode Decomposition and Machine Learning for Categorization of Cardiac Diseases*

Saeed Mian Qaisar (LINEACT CESI).

The arrhythmia is one of the cardiovascular diseases which has several types. In literature, researchers have presented a broad study on the strategies utilized for Electrocardiogram (ECG) signal investigation. Automated arrhythmia detection by analyzing the ECG data has been reported using a number of intriguing techniques and discoveries. In order to effectively categorize arrhythmia, a novel approach based on the hybridization of the denoising filter, QRS segmentation, "Empirical Mode decomposition" (EMD), "Intrinsic Mode Functions" (IMFs) based features extraction, and machine learning techniques is developed in this study. To evaluate the categorization accuracy, the 10-fold cross validation (10-CV) method is used. Using an arrhythmia dataset that is freely available, the performance of our method is evaluated. A 97 percent average accuracy rate in identifying arrhythmias which is a high score comparing with other approaches proposed in the literature.

[2735] *Evolutionary games in modeling cancer metastases*

Andrzej Swierniak (Silesian University of Technology), Katarzyna Hajdowska (Silesian University of Technology) and Damian Borys (Silesian University of Technology).

In our study we propose complex Evolutionary Game Theory (EGT) models of tumor-tumor cells interactions containing different strategies of dissemination of cancer which take into account results of clinical and medical imaging data. Moreover, we apply new tools of spatial evolutionary tools, proposed by us recently. These tools take into account heterogeneity at the cell level (the so called Mixed Spatial Evolutionary Games – MSEG) and varying in time (and possibly also in space) effects of environment (Evolutionary Games with Resources and Spatial Evolutionary Games with Resources, respectively). In the former case it leads to multilayer structure of the game and in the latter case to time varying pay-off tables. The main idea that autonomously growing cells because of evolutionary acquisition are able to become motile and invasive and afterwards, disseminate, first, to local and subsequently (or sometimes immediately), to distant sites leads to several questions which could be at least qualitatively answered by game theoretic model. First, we can ask what are factors deciding when and where distant metastases in a given patient will emerge. Based on patient characteristics and radiomics features from PET/CT scans such as metabolic tumor volume and total lesion glycolysis we may construct the pay-off table whose entries measure changes in evolutionary adjustment resulting from interaction of cells representing different phenotypes (division, motility, cell-cell contacts, apoptosis).

[2784] *Oral e-Health Monitoring Platform - a platform for sharing oral health data*

Mónica Fernandes (Universidade Católica Portuguesa, Faculty of Dental Medicine, Center for Interdisciplinary Research in Health (CIIS)), Eduardo Esteves (Universidade Católica Portuguesa, FMD, Center for Interdisciplinary Research in Health; PromptEquation ERGAP), Diogo Duarte (PromptEquation / ERGAP), Marlene Barros (Universidade Católica Portuguesa, Faculty of Dental Medicine, Center for Interdisciplinary Research in Health (CIIS)), Nuno Rosa (Universidade Católica Portuguesa, Faculty of Dental Medicine, Center for Interdisciplinary Research in Health (CIIS)) and André Correia (Universidade Católica Portuguesa, Faculty of Dental Medicine, Center for Interdisciplinary Research in Health (CIIS)).

While there are several clinical data management tools available in dentistry, there is currently a lack of solutions that empower patients to have control over their own data.

The Oral e-Health Monitoring Platform has been developed following extensive research and market studies, showcasing an innovative and pioneering approach that facilitates direct patient communication and support. The primary objective of this platform is to enhance the oral health of the population by centralizing and facilitating the seamless exchange of medical-dental clinical records between patients and dentists.

This platform offers a range of functionalities, including multi-language support, a hybrid application framework, direct dentist-patient communication, creation and sharing of clinical records with patients, clinical decision support based on decision trees, and an alert system for follow-up.

Looking ahead, our vision for the platform involves expanding its implementation to at least two additional countries, integrating with existing clinical record platforms in dentistry, and promoting its adoption within academic environments.

[2785] *Quality of Life Analysis of Patients with Dermatological Problems: Teledermatology Versus Face-to-Face Dermatology*

Antonio López-Villegas (Poniente University Hospital), Remedios López Liria (University of Almeria) and Maria Angeles Valverde-Martinez (University of Almeria).

Introduction: The health-related quality of life (HRQoL) of the patients cared for with teledermatology (TD) services was analyzed as compared with face-to-face dermatology (F-F/D) at the hospital. Methodology: This study was a controlled, non-blinded, intra-level, and multicenter randomized clinical trial, with a 6-month follow-up. A total of 450 patients were randomly assigned to two different groups. The Spanish version of the generic EuroQol-5-dimensions-5-Levels (EQ-5D-5L) questionnaire and the specific Skindex-29 questionnaire were used at 0 and 6 months. Results: The number of primary care visits (2.24 TD; 1.68 F-F/D) and number of hospital visits (0.01 TD; 1.48 F-F/D) were statistically signifi-cant. It was observed that from month 0 onwards, the users included in the F-F/D group self-perceived a lower HRQoL than the users included in the TD group (Skindex-29 total: $p \leq 0.00$; EQ-5D-5L VAS = $p \leq 0.00$; EQ-5D-5L utilities = $p \leq 0.00$). Conclusions: At the end of the study, the patients included in the F-F/D group still obtained lower scores in their perception of HRQoL, as compared to those included in the other type of follow-up (Skindex-29 total: $p \leq 0.00$; EQ-5D-5L VAS = $p \leq 0.00$; EQ-5D-5L utilities = $p \leq 0.00$). TD was an effective diagnosis and follow-up tool. At the end of the study period, the HRQoL of the patients in both groups was significantly higher as compared to their baseline levels. Additionally, both the general and specific HRQoL perceived by the TD patients was higher than the F-F/D group from the start of the study.

[2789] *Putative quadruplex forming sequences in plants: role, occurrence and possible interaction partners*

Adriana Volná (Department of Physics, Faculty of Science, University of Ostrava, Ostrava, Czech Republic), Martin Bartas (Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic) and Jiří Červeň (Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic).

It is well-known, that epigenetic regulatory mechanisms of gene expression are responsible for the transcription and translation control in plants. Although G-quadruplex forming structures are not usually attributed to the epigenetic regulatory mechanisms, they perfectly fit the definition - their formation can dramatically affect the replication, transcription, and even translation without alternating the genetic information. In this work, we used publicly available data in the databases and recently created tools to search for sequences with the potential to form G-quadruplexes (so-called putative quadruplex sites, PQSs), compared them among plant species, but also described their variable occurrence between different genomic features in model Pisum sativum. Another very interesting result of our research effort was the discovery of the highly conserved PQS in the RBP1 gene (coding for the RNA polymerase II large subunit) between very distant plant species which suggests their functional relevance. Finally, using bioinformatic predictive methods, we were able to propose several hundreds of candidate proteins (involved in diverse molecular and physiological processes in plants), which might not only interact but also regulate G-quadruplex formation, stability, and resolving in plant cells. Herein we intend to summarize acquired bioinformatical data into a coherent and clear hypothesis comprising dynamic regulation caused by the formation of these structures in plants including their interaction with protein partners which is worth testing also using in vitro techniques.

[2829] *Agent based modeling of fish shoal behavior*

Pavla Urbanova (University of South Bohemia in Ceske Budejovice), Ievgen Koliada (University of South Bohemia in Ceske Budejovice), Petr Cisar (University of South Bohemia in Ceske Budejovice) and Milos Zelezny (University of West Bohemia in Pilsen).

Fish require a sufficient amount of dissolved oxygen in the water to breathe and maintain their metabolic functions. Insufficient lev- els of dissolved oxygen can lead to stress, illness, and even death among the fish population. Therefore, it is crucial to model and simulate the relationship between dissolved oxygen levels and fish behavior in order to optimize aquarium design and management. One approach to studying this relationship is through multiagent based modeling. This method involves creating a virtual environment in which multiple agents, representing individual fish, interact with each other and the environment based on a set of predened rules. In the context of aquarium simulation, the agents would represent individual fish, and the environment would represent the aquarium water and its parameters.

[2884] *Recognition of conformational states of a G Protein-Coupled Receptor from molecular dynamic simulations using sampling techniques*

Mario Alberto Gutiérrez Mondragón (Universitat Politécnica de Catalunya, Computer Science Department, and IDEAI-UPC Research Center), Caroline König (Universitat Politécnica de Catalunya, Computer Science Department, and IDEAI-UPC Research Center) and Alfredo Vellido Alcacena (Universitat Politécnica de Catalunya, Computer Science Department, and IDEAI-UPC Research Center).

Protein structures are complex and dynamic entities relevant to many biological processes. G-protein-coupled receptors in particular are a functionally relevant family of cell membrane proteins of interest as targets in pharmacology. Nevertheless, the limited knowledge about their inherent dynamics hampers the understanding of the underlying functional mechanisms that could benefit rational drug design. The use of molecular dynamics simulations and their analysis using Machine Learning methods may assist the discovery of diverse molecular processes that would be otherwise beyond our reach. The current study builds on previous work aimed at uncovering relevant motifs (groups of residues) in the activation pathway of the β2-adrenergic (β2AR) receptor from molecular dynamics simulations, which was addressed as a multi-class classification problem using Deep Learning methods to discriminate active, intermediate, and inactive conformations. For this problem, the interpretability of the results is a relevant problem. Unfortunately, the vast amount of intermediate transformations, in contrast to the number of re-orderings establishing active and inactive conditions, handicaps the identification of relevant residues related to a conformational state as it generates a class-imbalance problem. The current study aims to investigate existing Deep Learning techniques for addressing such problem that negatively influences the results of the predictions, aiming to unveil a trustworthy interpretation of the information revealed by the models about the receptor functional mechanics.

[2926] *Entropy approach of processing for fish acoustic telemetry data to detect atypical behavior during welfare evaluation*

Jan Urban (Institute of Complex Systems, FFPW, USB).

Fish telemetry is an important tool for studying fish behavior, allowing to monitor fish movements in real-time. Analyzing telemetry data and translating it into meaningful indicators of fish welfare remains a challenge. This is where entropy approaches can provide valuable insights. Methods based on information theory can quantify the complexity and unpredictability of animal behavior distribution, providing a comprehensive understanding of the animal state. Entropy-based techniques can analyze telemetry data and detect changes in fish behavior, or irregularity. By analyzing the accelerometer data, using entropy approach, it is possible to identify atypical behavior that may be indicative of compromised welfare

---

[3046] *Deep Learning for automatic electroencephalographic signals classification*

Nadia N. Sánchez-Pozo (Universidad Politécnica Estatal del Carchi, Tulcán, 040101, Ecuador), Samuel Lascano-Rivera (Universidad Politécnica Estatal del Carchi, Tulcán, 040101, Ecuador), Francisco J. Montalvo-Marquez (Universidad Politécnica Estatal del Carchi, Tulcán, 040101, Ecuador) and Dalia Y. Ortiz-Reinoso (Universidad de Guayaquil, Guayaquil 090510, Ecuador).

Automated electroencephalographic (EEG) signals classification using deep learning algorithms is an emerging technique in neuroscience that has the potential to detect brain pathologies such as epilepsy efficiently. In this process, deep

learning algorithms are trained with labeled EEG signal datasets. However, due to the highly complex nature of EEG signals and the large amount of irrelevant information they contain, feature extraction techniques must be applied to reduce their dimension- ality and focus on relevant information. This paper presents a comparative study on feature extraction methods for the classification of EEG recordings. The results demonstrate that the proposed classification algorithms and characterisation techniques are effective and suitable, as the accuracy metrics reach a value of 99.27%. The results presented in this paper contribute to the further development of automatic EEG signal classification methods based on deep learning.

[3356] *Training Strategies for Covid-19 Severity Classification using Machine Learning and Heart Rate Variability Metrics*

Daniel Pordeus Menezes (FEDERAL UNIVERSITY OF CEARÁ), Pedro Ribeiro (Universidade Católica Portuguesa), Laíla Zacarias (Federal University of Ceará), João Madeiro (FEDERAL UNIVERSITY OF CEARÁ), João Marques (University of Saint Joseph), Pedro Rodrigues (Universidade Católica Portuguesa), Camila Leite (Federal University of Ceará), Manoel Neto (Federal University of Ceará), Arnaldo Peixoto Jr (Federal University of Ceará) and Adriel de Oliveira (University for the International Integration of the Afro-Brazilian Lusophony).

The COVID-19 pandemic has posed a significant public health challenge on a global scale. It is imperative that we continue to undertake research in order to identify early markers of disease progression, enhance patient care through prompt diagnosis, identification of high-risk patients, early prevention, and efficient allocation of medical resources. In this particular study, we obtained 100 5-minute electrocardiograms (ECGs) from 50 COVID-19 volunteers in two different positions, namely upright and supine, who were categorized as either moderately or critically ill. We used classification algorithms to analyze heart rate variability (HRV) metrics derived from the ECGs of the volunteers with the goal of predicting the severity of illness. Our study choose a configuration pro SVC that achieved 76% of accuracy, and 0.84 on F1 Score in predicting the severity of Covid-19 based on HRV metrics.

[3365] *The Utilization of Click Reactions in Understanding the Structure and Function of IgE in Allergic Reactions*

Parth Shinde (BITS Pilani K K Birla Goa Campus), Yash Saini (BITS Pilani K K Birla Goa Campus), Veeky Baths (BITS Pilani K K Birla Goa Campus), Manas Mandal (Roseman University) and Shounak Bhattacharya (BITS Pilani K K Birla Goa Campus).

This study examines cutting-edge techniques, namely click chemistry and antigen-antibody complex formations, in order to develop affordable and effective home allergy testing kits. The ultimate objective is to provide effective methods for diagnos- ing and treating allergic conditions. Through an azide click re- action, the surface of the hydrophilic polymer-coated paper is functionalized with captured antibodies and used as a diagnostic substrate for the detection of food allergies. After introducing a blood sample and a second antibody conjugated with a detec- tion moiety, the presence of specific IgE antibodies to the food allergen is determined by measuring the fluorescence or colour intensity of bound IgE. By comparing the test's results to inter- nationally recognized benchmarks, the test's accuracy is con- firmed. When examining how IgE interacts with antigens and its role in effector cell activation, the application of click chem- istry in this context has substantial implications for developing novel therapies for allergic diseases. Additional research in this field will increase our understanding of the underlying mecha- nisms, paving the way for future advances in allergy diagnosis and treatment.

[3517] *Application of custom approaches for target enrichment of SARS-CoV-2 in nasopharyngeal swab for whole genome sequencing*

Anna Gladkikh (Saint Petersburg Pasteur Institute), Ekaterina Klyuchnikova (Saint Petersburg Pasteur Institute), Valerya Sbarzaglia (Saint Petersburg Pasteur Institute), Dmitrii Polev (Saint Petersburg Pasteur Institute) and Vladimir Dedkov (Saint Petersburg Pasteur Institute).

Genomic surveillance plays a key role in overcoming the ongoing COVID-19 pandemic despite its relative successive waves and the continuous emergence of new variants. Two methods have been developed and implemented to enrich SARS-CoV-2 genomes in patient samples. One approach is PCR enrichment with the author's panel of primers. The second approach is enrichment with probes suitable for targets that can be amplified by PCR. Both approaches have shown their

effectiveness and ease of use for obtaining genome-wide sequences of SARS-CoV-2 and are successfully used in the monitoring of SARS-CoV-2 genovariants in the northwestern region of the Russian Federation.

[3696] *Clinical text classification in Cancer Real-World Data in Spanish*

Francisco J. Moreno-Barea (Universidad de Málaga), Héctor Mesa (Universidad de Málaga), Nuria Ribelles (Unidad de Gestión Clínica Intercentros de Oncología de Málaga), Emilio Alba (Unidad de Gestión Clínica Intercentros de Oncología de Málaga) and Jose M. Jerez (Universidad de Málaga).

Healthcare systems currently store a large amount of clinical data, mostly unstructured textual information, such as electronic health records (EHRs). Manually extracting valuable information from these documents is costly for healthcare professionals. For example, when a patient first arrives at an oncology clinical analysis unit, clinical staff must extract information about the type of neoplasm in order to assign the appropriate clinical specialist. Automating this task is equivalent to text classification in natural language processing (NLP). In this study, we have attempted to extract the neoplasm type by processing Spanish clinical documents. A private corpus of 23,704 real clinical cases has been processed to extract the three most common types of neoplasms in the Spanish territory: breast, lung and colorectal neoplasms. We have developed methodologies based on state-of-the-art text classification task, strategies based on machine learning and bag-of-words, based on embedding models in a supervised task, and based on bidirectional recurrent neural networks with convolutional layers (C-BiRNN). The results obtained show that the application of NLP methods is extremely helpful in performing the task of neoplasm type extraction. In particular, the 2-BiGRU model with convolutional layer and pre-trained fastText embedding obtained the best performance, with a macro-average, more representative than the micro-average due to the unbalanced data, of 0.981 for precision, 0.984 for recall and 0.982 for F1-score.

[3767] *Systematic comparison of advanced network analysis and visualization of lipidomics data*

Jana Schwarzerová (Department of Biomedical Engineering, BUT, Czechia; Molecular Systems Biology, University of Vienna, Austria), Dominika Olešová (Institute of Experimental Endocrinology, & Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovak Republic), Aleš Kvasnička (Department of Clinical Biochemistry, Palacký University Olomouc and University Hospital Olomouc, Czech Republic), David Friedecký (Department of Clinical Biochemistry, Palacký University Olomouc and University Hospital Olomouc, Czech Republic), Margaret Varga (Department of Biology, University of Oxford, Oxfordshire), Valentine Provazník (Department of Biomedical Engineering, Brno University of Technology, Technicka 12, 616 00 Brno, Czech Republic) and Wolfram Weckwerth (Molecular Systems Biology (MOSYS); Vienna Metabolomics Center (VIME), University of Vienna, Vienna, Austria).

Comprehensive analysis of lipids is becoming a forefront of clinical data analysis. Due to significant technical advancements, lipidomics is emerging in clinical diagnostics for improvement and earlier detection of a broad range of diseases. However, in order to understand the biological complexities and interrelationships between the molecules, it is important to have a correct representation of the data and visualizations that enable good interpretability of the lipidomic data. Therefore, the present study systematically compares different visualization methods for lipidomic data, based on different com-putational relations between the selected lipids and supplemented with known biological information. Networks were reconstructed, and an analysis was performed to objectively compare the visualizations.

[3798] *Exploring machine learning algorithms and protein language models strategies to develop enzyme classification systems*

Diego Fernández (Departamento de Ingeniería en Computación, Universidad de Magallanes), Álvaro Olivera-Nappa (Departamento de Ingeniería Química, Biotecnología y Materiales, Universidad de Chile), Roberto Uribe-Paredes (Departamento de Ingeniería en Computación, Universidad de Magallanes) and David Medina (Departamento de Ingeniería en Computación, Universidad de Magallanes).

Discovering functionalities for unknown enzymes has been one of the most common bioinformatics tasks. Functional annotation methods based on phylogenetic properties have been the gold standard in every genome annotation process. However, these methods only succeed if the minimum requirements for expressing similarity or homology are met. Alternatively, machine learning and deep learning methods have proven helpful in this problem, developing functional

classification systems in various bioinformatics tasks. Nevertheless, there needs to be a clear strategy for elaborating predictive models and how amino acid sequences should be represented. In this work, we address the problem of functional classification of enzyme sequences (EC number) via machine learning methods, exploring various alternatives for training predictive models and numerical representation methods. The results show that the best performances are achieved by applying representations based on pre-trained models. However, there needs to be a clear strategy to train models. Therefore, when exploring several alternatives, it is observed that the methods based on CNN architectures proposed in this work present a more outstanding facility for learning and pattern extraction in complex systems, achieving performances above 97% and with error rates lower than 0.05 of binary cross entropy. Finally, we discuss the strategies explored and analyze future work to develop integrated methods for functional classification and the discovery of new enzymes to support current bioinformatics tools.

[3869] *BodyFlow: A library for human pose estimation and activity recognition*

Irene López-Bosque (ITAINNOVA), Carlos Marañes-Nueno (ITAINNOVA), Ana Caren Hernández-Ruiz (ITAINNOVA), Rocío Aznar-Gimeno (ITAINNOVA), Pilar Salvo-Ibañez (ITAINNOVA), María de La Vega Rodrigálvarez-Chamarro (ITAINNOVA), David Abadía-Gallego (ITAINNOVA) and Rafael del-Hoyo-Alonso (ITAINNOVA).

Human activity recognition is considered an essential task for various applications in healthcare, sports, security, gaming, and other fields. The proposed library allows the identification of common activities based on the human 2D/3D pose estimated from visual data. Furthermore, the library is capable of working simultaneously with inertial sensor information, granting the user the flexibility to select the input data of their choice. BodyFlow comprises seven state-of-the-art algorithms for pose estimation and three different neural networks for human activity recognition.

[4049] *Improving Foetal Health Monitoring: A Review of the Latest Developments and Future Directions*

Restuning Widiasih (Faculty of Nursing, Universitas Padjadjaran, Bandung, Indonesia), Hasballah Zakaria (School of Electrical Engineering and Informatics Institut Teknologi Bandung, Bandung, Indonesia) and Siti Saidah Nasution (Faculty of Nursing Universitas Sumatra Utara, Medan, Indonesia).

Devices for monitoring heart rate and fetal movement are becoming increasingly sophisticated with the latest technological advancements. However, there is a pressing need for a more comprehensive review and analysis of these tools. Aim, the objective of this literature review is to identify fetal health monitoring devices, evaluate the sensitivity of monitoring fetal heart rate and growth/movement, and determine the target users for these devices. Method, the search was conducted using PubMed and Scopus databases, with PICO-based keywords that included pregnant women or pregnancy as the population, fetal or heart rate monitoring tool or fetal movement tool with sensitivity or reactivity as the research interest. A total of 2077 papers were initially identified, with 36 selected after article screening using the PRISMA approach and critical appraisal with JBI. Results, the analysis classified the devices into three categories: Fetal Heart Monitoring, Fetal Movement, and Fetal Heart Monitoring and Movement. The monitoring technology applied wave detection through cardiography and myography. The devices demonstrated a signal sensitivity of more than 75% for both the mother and the fetus. Conclusions, the analysis of the 36 articles revealed that monitoring technology is rapidly evolving, but almost all devices are designed for use by health workers. Only the Piezo Polymer Pressure Sensor is intended for independent monitoring by mothers and families. The development and research of independent fetal monitoring are necessary to improve monitoring during the new adaptation period after the Covid-19 pandemic.

[4092] *A System Biology and Bioinformatics approach to determine the molecular signature, core ontologies, functional pathways, drug compounds in between Stress and Type 2 Diabetes*

Md. Abul Basar (Department of ICT, Mawlana Bhashani Science and Technology University, Bangladesh), Md. Rakibul Hasan (Department of Software Engineering (SWE), Daffodil International University (DIU), Ashulia, Dhaka, Bangladesh), Bikash Kumar Paul (Department of ICT, Mawlana Bhashani Science and Technology University, Bangladesh), Khairul Alam Shadhin (Department of Software Engineering (SWE), Daffodil International University (DIU), Ashulia, Dhaka, Bangladesh) and Md. Sarwar Mollah (Department of CIS, Daffodil International University (DIU), Ashulia, Dhaka, Bangladesh).

Bioinformatics is the application of computer science and information technology to the field of biology and medicine. It involves the analysis of large amounts of biological data, such as DNA sequences, protein structures, and gene expression patterns. Bioinformatics is used to develop new methods for understanding and analyzing biological data, as well as to develop new tools and technologies for biological research. Bioinformatics is used in a variety of fields, including genomics, proteomics, and drug discovery. In this study, focus on two severe diseases which affect millions of people globally such as stress and type 2 diabetes. Stress can have a significant impact on people with type 2 diabetes. Stress can cause blood sugar levels to rise, making it difficult to manage diabetes. The purpose of this research is to use various bioinformatics methods to discover potential therapeutic drugs and functional pathways between stress and type 2 diabetes. The microarray datasets GSE183648 and GSE20966 are used for the analysis of stress and type 2 diabetes samples respectively. After the datasets have been preprocessed and filtered through the use of the R programming language, identified the common DEGs. The depiction of common DEGs is shown by venn diagram. Next, the most active genes are identified through topological properties, and PPIs are built from the similar differential expressed genes (DEGs). These five genes NTRK2, SOCS3, NEDD9, MAP3K8, and SIRPA are the most important hub genes with in the interaction network of protein-protein. According to the common DEGs, GO term's molecular function (MF), KEGG and WikiPathways are shown in this study. Gene-miRNA interaction, TF-gene regulatory network, module analysis, GO terms (Biological Process, Cellular Component), Pathways (Reactome, BioCarta, BioPlanet) are all things that could be done with this research work in the future. In last, a therapeutic drug compounds are recommended on the basis of common DEGs.

[4206] *Comparison of image processing and classification methods for a better diet decision-making*

Maryam Abbasi (Institute Polytechnic of Coimbra), Pedro Martins (Polytechnic Institute of Viseu, Dep. of Computer Sciences) and Filipe Cardoso (Polytechnic Institute of Viseu, Dep. of Computer Sciences).

This paper aims to explore the use of different deep learning techniques, specifically convolutional neural networks (CNNs), for dietary assessment through image food recognition and compare their performance to the human visual system (HVS). Currently, there are three main techniques for using CNNs in this task: training a network from scratch; using an off-the-shelf pre-trained network; and performing unsupervised pre-training with supervised adjustments. In this study, the authors evaluate the performance of three CNN models with varying numbers of parameters (5,000 to 160 million) based on dataset size and spatial image context.

The authors also consider human knowledge and classification to compare the performance of the CNNs to the HVS. They find that while the CNNs make errors across different food classes, the HVS tends to make semantic errors with specific food classes. As a result, the HVS shows more consistency in its answers. Overall, the findings suggest that the HVS is more accurate when the dataset is diverse, while the CNN performs better when the dataset is focused on a particular niche.

In conclusion, this study provides empirical evidence that machine learning can be more efficient than the HVS in certain tasks but also highlights the strengths and limitations of both approaches. The authors suggest that combining CNNs with other classification techniques, such as bag-of-words, may be a promising approach for improving the accuracy of dietary assessment through image food recognition.

[4218] *EMCNN: Fine-Grained Emotion Recognition based on PPG using Multi-scale Convolutional Neural Network*

Jiyang Han (Northwestern Polytechnical University) and Hui Yang (Northwestern Polytechnical University).

Automatic emotion recognition based on physiological signals has been widely studied in the recent years due to the superiority of difficult to pretend. However, most of the methods are based on multi-channel or multi-mode signals such as EEG which are difficult to obtained during the daily life, leading to application limitation. Thanks to easy acquirement, photoelectric plethysmograph (PPG) is more promising for automatic emotion analysis. As a metric of hemodynamics, it can reflect the emotion change indirectly. However, PPG is a single channel signal with less discriminative information for emotion, making poor accuracy on fine-grained recognition. To solve this problem, we proposed a Multi-scale Convolutional Neural Network for Emotion recognition called EMCNN. The proposed EMCNN can be further divided into three stages: transformation stage, feature extraction stage and classification stage. In the first stage, the data dimension can be increased by using transformation method at the time domain and the frequency domain on the meta data separately. In this way, the discriminative information in the PPG signal is augmented accompanied by the dimension increment. In the second stage, depthwise separable convolution is introduced to extract features of each dimension separately. Finally, features from all of dimensions are concreted to give a final classification result. EMCNN was

evaluated on a widely-used public dataset and a self-collected dataset. Experiments results on the DEAP dataset shows that our proposed EMCNN can achieve the overall accuracy of 78% for 2-classe recognition and 67% for 4-class recognition which can be comparable to most of the state-of-art methods based on other physiological signal signals. Moreover, our model can achieve an acceptable performance on the self-dataset to demonstrate its generality. In this work, we first utilize PPG signal to achieve fine-grained automatic emotion recognition. Furthermore, the proposed EMCNN is a promising method to solve the single channel signal analysis.

[4285] *Pharmacoinformatics Analysis of Drug Leads for Alzheimer's Disease from FDA-Approved Dataset through Drug Repositioning Studies*

Mubashir Hassan (Nationwide Children's Hospital), Saba Shahzadi (Nationwide Children's Hospital) and Andrzej Kloczkowski (Nationwide Children's Hospital).

Rational drug designing is long lasting, time consuming and highly priced approach with particularly low success rate. To overcome this problem, computational drug repositioning approach is the most effective approach to predict the possible alternative therapeutic effects of already FDA approved drugs against different diseases. In this study, we had designed a computational mechanistic approach to fetch the promising drugs from the pool of FDA approved dataset against Alzheimer's Disease (AD) based on combination of shape-based screening, molecular docking, pharmacogenomics and MD simulation studies, respectively. The binding interaction patterns and conformations of screened drugs within active region of MAP kinase-activating death domain protein (MADD) were confirmed through molecular docking profiles. The possible associations of selected drugs with AD genes were predicted by pharmacogenomics analysis and confirmed through data mining. The stability behavior of the docked complexes was evaluated by MD simulations through analyses of root-mean-square deviations/fluctuations (RMSD/F), radius of gyration (Rg) and solvent-accessible surface area (SASA) graphs. Taken together, darifenacin, astemizole, tubocurarine, elacridar, sertindole and tariquidar displayed best results and might be used as possible therapeutic agents in the treatment of AD after future de-tailed in vitro and in vivo assessments.

[4314] *Determining HPV Status in Patients with Oropharyngeal Cancer from 3D CT Images Using Radiomics: Effect of Sampling Methods*

Kubra Sarac (Biomedical Engineering, Bogazici University) and Albert Guvenis (Biomedical Engineering, Bogazici University).

Non-invasive detection of human papillomavirus (HPV) status is important for the treatment planning of patients with oropharyngeal cancer (OPC). In this work, three-dimensional (3D) head and neck computed tomography (CT) scans are utilized to identify HPV infection status in patients with OPC by applying radiomics and several resampling methods to handle highly imbalanced data. 1142 radiomic features were obtained from the segmented CT images of 238 patients. The features used were selected through correlation coefficient analysis, feature importance analysis, and backward elimination. The fifty most important features were chosen. Six different sampling methods which are Synthetic Minority Oversampling Technique (SMOTE), Support Vector Machine Synthetic Minority Oversampling Technique (SVMSMOTE), Adaptive Synthetic Sampling Method (ADASYN), NearMiss, Condensed Nearest Neighbors (CNN), and Tomek's Link were performed on the training set for each of the positive and negative HPV classes. Two different machine learning (ML) algorithms a Light Gradient Boosting Machine (LightGBM) and Extreme Gradient Boosting (XGBoost) were applied as predictive classification models. Model performances were assessed separately on 20% of the data. Oversampling methods displayed better performance than undersampling methods. The best performance was seen in the combination of SMOTE and XGBoost algorithms which had an area under the curve (AUC) of 0.93 (95% CI: 82-99) and an accuracy of 90% (95% CI: 78-96). Our work demonstrated a reasonable accuracy in the forecast of HPV status by using 3D imbalanced and small datasets by using resampling methods. Further work is needed to test the algorithms on larger, balanced, and multi-institutional data.

[4398] *Multiallelic Maximal Perfect Haplotype Blocks with Wildcards via PBWT*

Paola Bonizzoni (DISCo, Univ. degli Studi Milano-Bicocca), Gianluca Della Vedova (DISCo, Univ. degli Studi Milano-Bicocca), Yuri Pirola (DISCo, Univ. degli Studi di Milano-Bicocca), Raffaella Rizzi (DISCo, Univ. degli Studi Milano-Bicocca) and Mattia Sgrò (DISCo, Univ. degli Studi Milano-Bicocca).

Computing maximal perfect blocks of a given panel of haplotypes is a crucial task for efficiently solving problems such as polyploid haplotype reconstruction and finding identical-by-descent segments shared among individuals of a population. Unfortunately, the presence of missing data in the haplotype panel limits the usefulness of the notion of perfect blocks. We propose a novel algorithm for computing maximal blocks in a panel with missing data (represented as wildcards). The algorithm is based on the Positional Burrows-Wheeler Transform (PBWT) and has been implemented in the tool Wild-pBWT, available at https://github.com/AlgoLab/Wild-pBWT/. Experimental comparison showed that Wild- pBWT is 10-15 times faster than another state-of-the-art approach, while using a negligible amount of memory.

[4486] *Medical X-ray Image Classification Method Based on Convolutional Neural Networks*

Veska Gancheva (Technical University of Sofia) and Tsviatko Jongov (Technical University of Sofia).

Artificial intelligence and machine learning, including convolutional neural net-works are increasingly entering the field of healthcare and medicine. The aim of the study is to optimize the learning process of convolutional neural networks through X-ray images preprocessing. A model for optimizing the overall architecture of a classifying convolutional neural network of chest X-rays by reducing the total number of convolutional operations is presented. The experimental results prove the successful application of the optimization process on the training of classification convolutional networks. There is a significant reduction in the training time of each epoch in the optimized convolutional networks. The optimization is of the order of 25% for the network with an input layer size of 124 x 124 and about 27% for the network with an input layer size of 122 x 122. The method can be applied in any field of image classification in which the informative image regions are grouped and subject to segmentation.

[4664] *Relation Predictions in Comorbid Disease Centric Knowledge Graph Using Heterogeneous GNN Models*

Saikat Biswas (Indian Institute of Technology, Kharagpur), Koushiki Dasgupta Chaudhuri (Indian Institute of Technology, Kharagpur), Pabitra Mitra (Indian Institute of Technology, Kharagpur) and Krothapalli Sreenivasa Rao (Indian Institute of Technology, Kharagpur).

Disease comorbidity has been an important topic of research for the last decade. This topic has become more popular due to the recent outbreak of COVID-19 disease. A comorbid condition due to multiple concurrent diseases is more fatal than a single disease. These comorbid conditions can be caused due to different genetic as well as drug-related side effects on an individual. There are already successful methods for predicting comorbid disease associations. This disease-associated genetic or drug-invasive information can help infer more target factors that cause common diseases. This may further help find out effective drugs for treating a pair of concurrent diseases. In addition to that, the common drug side-effects causing a disease phenotype and the gene associated with that can be helpful in finding important biomarkers for further prognosis of the comorbid disease. In this paper, we use the knowledge graph (KG) from our previous study to find out target-specific relations apart from sole disease-disease associations. We use four different heterogeneous graph neural network models to perform link prediction among different entities in the knowledge graph and we perform a comparative analysis among them. It is found that our best heterogeneous GNN model outperforms existing state-of-the-art models on a few target-specific relationships. Further, we also predict a few novel drug-disease, drug-phenotype, disease-phenotype, and gene-phenotype associations. These interrelated associations are further used to find out the common phenotypes associated with a comorbid disease as well as caused by the direct side effects of a treating drug. In this regard, our methodology also predicts some novel biomarkers and therapeutics for different fatal prevalent diseases.

[4683] *Prediction of Functional Effects of Protein Amino Acid Mutations*

Óscar Álvarez-Machancoses (University of Oviedo), Eshel Faraggi (Indiana University Purdue University, Indianapolis), Enrique Deandrés-Galiana (University of Oviedo), Juan Fernández-Martínez (University of Oviedo) and Andrzej Kloczkowski (Nationwide Children Hospital).

Human Single Amino Acid Polymorphisms (SAPs) or Single Amino Acid Variants (SAVs) usually named as nonsynonymous Single Nucleotide Variants nsSNVs) represent the most frequent type of genetic variation among the population. They originate from non-synonymous single nucleotide variations (missense variants) where a single base pair substitution alters the genetic code in such a way that it produces a different amino acid at a given position. Since mutations are commonly associated with the development of various genetic diseases, it is of utmost importance to understand and predict which variations are deleterious and which are neutral. Computational tools based on machine learning are

becoming promising alternatives to tedious and highly costly mutagenic experiments. Generally, varying quality, in-completeness and inconsistencies of nsSNVs datasets degrade the usefulness of machine learning approaches. Consequently, robust and more accurate approaches are essential to address these issues. In this paper, we present the application of a consensus classifier based on the holdout sampling, which shows robust and accurate results, outperforming currently available tools. We generated 100 holdouts to sample different classifiers' architectures and different classification variables during the training stage. The best performing holdouts were selected to construct a consensus classifier and tested by blindly utilizing a k-fold $(1 \leq k \leq 5)$ cross-validation approach. We also per-formed an analysis of the best protein attributes for predicting the effects of nsSNVs by calculating their discriminatory power. Our results show that our method outperforms other currently available tools, and provides robust results, with small standard deviations among folds and high accuracy. The superiority of our algorithm is based on the utilization of a tree of holdouts, where different machine learning algorithms are sampled with different boundary conditions or different predictive attributes.

[4804] *Computational study of conformational changes in intrinsically disordered regions during protein-protein complex formation*

Amita Barik (National Institute of Technology, Durgapur), Madhabendra Mohon Kar (National Institute of Technology, Durgapur) and Prachi Bhargava (National Institute of Technology, Durgapur).

A significant section of the proteome of an organism, especially that of eukaryotes consists of polypeptide segments that are functional, even though they cannot form a defined three-dimensional structure and rather exists as an ensemble of conformations. These are known as Intrinsically Disordered Regions (IDRs). Proteins that lack IDRs are called structural or ordered proteins, while those composed entirely of disordered regions are called intrinsically disordered proteins (IDPs). Most proteins are a combination of IDRs and structured regions. It has been reported that IDRs play a pivotal role in modulating cellular processes and signaling pathways. Furthermore, several studies suggest that IDRs can provide valuable insights into drug design as their structural disorder can aid in ligand selection for drug development. Therefore, it is crucial to understand the structural and functional properties of IDRs and their effect on macromolecular interactions. IDRs often act as hubs in protein-protein interaction (PPI) networks due to their conformational flexibility, allowing them to interact with multiple biomolecules. They can adopt a well-defined tertiary structure upon binding, retain some degree of disorder, or remain completely disordered, forming fuzzy complexes. In the present study, we analyse the conformational changes in IDRs upon complex formation using a non-redundant dataset of binary, X-ray solved protein-protein (P-P) complexes and their corresponding unbound forms. IDRs are prevalent in the unbound forms of the complexes: in 41 unbound forms, 52 IDRs were found. The length of these IDRs ranges from 2 to 82 residues and they are depleted in aromatic amino acids (1.2%). We study the transitions from disordered to ordered regions during complex formation and categorise them into three classes: disordered in unbound and ordered in complex (D-O), disordered in unbound and disordered in complex (D-D) and disordered in unbound and partially ordered in complex (D-PO). We find 10%, 62%, and 28% of the IDRs present in the unbound proteins in the D-O, D-D and D-PO class respectively. The study of secondary structures of residues in D-O category reveal that 79.78% form coils indicating that IDRs in proteins tend to form ordered coils upon binding to suitable partners. We also observe that 42.9% residues under D-O category are located at the interface of P-P complexes suggesting that they contribute to the stability of the complexes. Amino acids under the D-O category also contribute in Hydrogen bond formation (0.7%) in the P-P complexes. Our findings provide fundamental insights into the underlying principles of molecular recognitions by disordered regions. There are some structured regions in the unbound proteins which upon complexation become disordered and, in our dataset, we observe 34.2% of such IDRs in the P-P complexes. It will be intriguing to learn more about these ordered-to-disordered transitions upon complex formation.

[4821] *WATER DYNAMICS IN CHEESE BY MEANS OF NUCLEAR MAGNETIC RESONANCE RELAXOMETRY*

Monika Małkowska-Kowalczyk (University of Warmia and Mazury in Olsztyn), Justyna Żulewska (University of Warmia and Mazury in Olsztyn), Adriana Lobacz (University of Warmia and Mazury in Olsztyn), Maciej Maciejczyk (University of Warmia and Mazury in Olsztyn) and Danuta Kruk (University of Warmia and Mazury in Olsztyn).

Nuclear Magnetic Resonance (NMR) relaxometry has been exploited to enquire into dynamical properties of water molecules entrapped in macromolecular matrices of cheese. NMR relaxometry is an unique method highly appreciated in molecular science. The great advantage of NMR relaxometry is the opportunity to probe dynamical processes over a broad time scale from milliseconds to nanoseconds and reveal the mechanism of the molecular motion. The potential of

NMR relaxometry results from the fact that in contrast to classical NMR experiments performed at a single magnetic field, NMR relaxometry covers a range of magnetic field (and hence resonance frequency) encompassing four orders of magnitude (from 10 kHz to 40 MHz referring to 1H resonance frequency). A set of cheeses of different macroscopic (physico-chemical) properties was selected: mozzarella (n=9), cream cheese (n=9), low moisture mozzarella cheese (LMMC) (n=8), LMMC with milk fat replaced by plant oil (n=3), low-fat cheese (n=3) and long ripening cheese (n=3). In each group, the cheeses from different production plants and different batches were used. The composition and water activity was measured using FoodScan analyzer (type 78810 FOSS, Denmark) and AQUA LAB 4TEV analyzer (type S40001855, USA), respectively. 1H spin-lattice relaxation rates was measured for the cheese samples in the frequency range from 10kHz to 10MHz at room temperature. The data were analyzed in terms of models linking the relaxation features with parameters directly characterizing the motion and organization of water molecules. As a results of such analyses the profiles, so-called fingerprints, specific to given groups of cheeses were created. The analyzes showed that the water activity did not correspond directly to the water dynamics presented by NMR profiles. However, it was stated that at lower resonance frequency values, the lower water content corresponds the higher level of relaxation.

[4874] *Evaluation of homogeneity of effervescent tablets containing quercetin and calcium using X-ray microtomography and hyperspectral analysis*

Michał Meisner (Department of Basic Biomedical Science, Faculty of Pharmaceutical Sciences, Medical University of Silesia in Katowice), Piotr Duda (Institute of Biomedical Engineering, Faculty of Science and Technology, University of Silesia), Beata Szulc-Musioł (Department of Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, Medical University of Silesia in Katowice) and Beata Sarecka-Hujar (Department of Basic Biomedical Science, Faculty of Pharmaceutical Sciences, Medical University of Silesian in Katowice).

Background: Drug stability describes its ability to maintain physical, chemical, therapeutic, and microbiological properties during storage. In the study present study, we aimed to assess where there are any differences in the homogeneity of the effervescent tablets containing quercetin and calcium using hyperspectral imaging and X-ray microtomography. Effervescent tablets show the combination of the advantages of a solid and a liquid form of the active substance [1,2]. We selected tablets containing calcium and quercetin as a model preparation. Quercetin is one of the most common naturally occurring polyphenols. Material and Methods: The effervescent Alercal tablets were selected as a model preparation in the present analysis (Zdrovit; Natur Produkt Pharma sp. z o. o., Ostrów Mazowiecka, Poland). We analyzed unexpired, expired, as well as stressed effervescent tablets which were stored at 40oC for 14 days. The research was funded by Medical University of Silesia in Katowice, Poland (project number PCN-1-058/K/2/O). Tablets were carefully evaluated in terms of external appearance, weight, diameter, thickness, disintegration time, and hardness according to Polish Pharmacopeia [3]. A hyperspectral analysis within the range from 400 nm to 1030 nm was performed using a Specim IQ hyperspectral camera (Spectral Imaging Ltd., Oulu, Finland). An X-ray microtomography analysis was performed using Phoenix v|tome|x scanner (GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany). Statistical analysis was performed using Statistica 13.0 software (StatSoft; Statistica, Tulsa, OK, USA). Results: Both unexpired and expired tablets met the pharmacopeial requirements. The homogeneity analysis showed significant differences between the three types of tablets. In turn, significantly higher reflectance values in both the visible and near-infrared range were demonstrated in unexpired tablets compared to expired and stressed tablets ($p=0.001$ and $p<0.001$, respectively). In the microtomographic analysis, the calibration phantom containing the areas with known reference density was scanned together with the analyzed effervescent tablets. The calibration curve of the correlation between the brightness of the pixels and the density of the standards from the calibration phantom was characterized by the equation $y=181.08x-166.09$. This analysis showed that the densities of unexpired (1.435 g/cm3) and stressed tablets (1.433 g/cm3) were significantly higher compared to the density of expired tablets and respectively (1.414 g/cm3, $p<0.001$). Thus, unexpired and stressed tablets had better homogeneity than expired tablets. Conclusions: The changes in tablets' homogeneity may indicate possible physical changes in effervescent tablets during storage, especially in conditions deviating from those recommended by the manufacturer.

References 1. Herlina, Kuswardhani, N., Belgis, M., Tiara, A. Characterization of Physical and Chemical Properties of Effervescent Tablets Temulawak (Curcuma zanthorrhiza) in the Vari-ous Proportion of Sodium Bicarbonate and Tartaric Acid. E3S Web of Conferences 142, 03006 (2020) 2. Patel, S.G., Siddaiah, M. Formulation and evaluation of effervescent tablets: a review. J Drug Deliv Ther. 8(6), 296-303 (2018). 3. Polish Pharmacopoeia, 12th ed., The President of the Office for Registration of Medicinal Products, Medical Devices and Biocidal Products, Warsaw (2020)

[4883] *A 20-year journey of tracing the development of web catalogues for rare diseases*

João Almeida (University of Aveiro) and José Luis Oliveira (University of Aveiro).

Rare diseases are affecting over 350 million individuals on a worldwide scale. However, studying such diseases is challenging due to the lack of individuals compliant with the study protocols. This unavailability of information raises some challenges when defining the best treatments or diagnosing patients in the early stages. Multiple organizations invested in sharing data and resources without violating patient privacy, which resulted in several platforms focused on aggregating information. Despite the benefits of these solutions, the evolution of data regulations leads to new challenges that may not be fully addressed in such platforms. Therefore, in this paper, we proposed an enhanced version of one of the identified open-source platforms for this purpose. With this work, we were able to propose different strategies for aggregating and sharing information about rare diseases, as well as to analyse the technological evolution when producing tools for biomedical data sharing, namely by analysing the evolution of the selected tool over the last two decades.

[4885] *MODULATION OF CYP1A1 AND CYP1B1 GENE EXPRESSION IN 3-METHYLCHOLANTHRENE-INDUCED PROSTATE CANCER BY AFRICAN PEAR FRUIT (DACRYODES EDULIS) IN WISTAR RATS.*

David Omisore (Student of New Vision University, Georgia.).

Prostate cancer is one of the most frequently diagnosed cancers in men. The African pear fruit and its oil has been reported to possess various medicinal properties. This study was carried out to determine the effects of the pulp oil of the African pear fruit on the expression of the phase one enzymes, CYP1A1 and CYP1B1, in 3-methylcholanthrene induced prostate carcinogenesis. The pulp oil was extracted using both soxhlet and cold extraction methods, with n-hexane and methanol. 21 days old male Wistar rats (34) were used in this study and divided into 4 groups. Groups B and D were fed with a diet containing the pulp oil (10%). Groups C and D intraperitoneally received 3-methylcholanthrene (200mg/kg body weight) after 21 days of feeding with normal rat pellet, while group A served as the control group. Results indicated that CYP1A1 was significantly expressed ($p < 0.05$) in group D (pre-treated group) when compared to other groups. The CYP1B1 was also significantly expressed ($p < 0.05$) in the group D when compared with the other groups, group C (negative control group) showed no significant difference. The mean body weight for each group increased throughout the weeks, with group C (171.8±54.33) with the highest mean weight when compared to other groups. The tumor weight and volume of group D (0.12±0.04; 0.08±0.03) was lower than group C (0.19±0.02; 0.11±0.01). Histopathology results reveal the mitigating effect of the pulp oil on the liver of the rats in group B (positive control) and group D when compared to that of group C showing severe hepatic steatosis, moderately infiltrated cytoplasms, and mild infiltration of inflammatory cells of the cells of the sinusoids. The testes tissues of the rats in group B and group D showed improved health status when compared to group C showing atrophic seminiferous tubules exhibiting thickened propria, and tubules reflecting maturation arrest. This study therefore showed that the pulp oil of the African pear fruit to be effective against 3-methylcholanthrene-induced prostate carcinogenesis.

[4962] *Effect of the primary macroelements in the susceptibility to diseases: A Review*

Grethel Lázara Sieiro Miranda (Sugarcane Research Institute), Alberto Nicolás González Marrero (Sugarcane Research Institute), Eida Luisa Rodríguez Lema (Sugarcane Research Instituto) and Mérida Rodríguez Regal (Sugarcane Research Institute).

The nutrition of the plant has an important function in the appearance and susceptibility of the crop. The present work intends an analysis of the state of the art regarding of the effect of the mineral nutrition, in particular the primary macroelements about the susceptibility to diseases. The revision behaved by means of the study of original scienti c articles and bibliographical revisions, published in Cuban and foreigners; as well as the works carried out as thesis in universities of the region and of the world. Although there are some contradictions, the primary macroelements can affect tolerance and / or resistance to plant diseases caused by various pathogens. The most observed effect is the decrease in the symptoms of the disease except for the nitrogen, the one which in excess or de cit leads to higher incidence in the increase of susceptibility. An appropriate supply of the fertilization that promotes the good growth of the plants is an element to consider for the resistance to diseases.

[5017] *Modeling and Simulation of Multiphase Flow in Integrated Multitrophic Aquaculture Systems with Macroalgae: Application of CFD-DEM*

Radomir Filip (CTU in Prague, Faculty of Mechanical Engineering, Department of Process Engineering, Prague), Pavla Urbanova (University of South Bohemia in České Budějovice, FFPW, ICS), Ingrid Masalo (UPC BarcelonaTECH, Departament d'Enginyeria Agroalimentària i Biotecnologia, Castelldefels) and Stepan Papacek (University of South Bohemia, FFPW, ICS Nove Hrady).

To better predict the most important features governing the macroalgae growth and nutrient removal in integrated multitrophic aquaculture systems, computational techniques can be applied for the numerical assessment of the key parameters that influence the biotechnological process. Recent advances in both computational hardware and software, such as the discrete element method (DEM) coupled with computational fluid dynamics (CFD) codes, have enabled simulations of multiphase flow in biotechnological systems. In this work, we perform CFD–DEM simulations of macroalgae motion within integrated multitrophic aquaculture (IMTA) systems, making one step towards modeling and CFD simulation of multiphase flow in tanks with macroalgae, i.e. we study the fluid-structure interaction problem in bubbled tanks with a highly flexible solid phase. Based on the experimental identification of material and geometrical properties of macroalgae Ulva sp., we run a set of detailed 2-D CFD studies for two types of bioreactor settings differing in the presence or absence of an inner cylinder assembly. Consequently, corresponding regression models for the macroalgae cyclic motion are derived, and the initial hypothesis consisting of the beneficial role of the assembly is confirmed. Eventually, the CFD results for a limited number of operating conditions are verified by an image processing technique on the laboratory scale tank, as well as by the 3-D CFD-DEM simulations.

[5108] *Bioinformatics for Transcription Factor Discovery at a Primarily Undergraduate Institution*

Michael Van Dyke (Kennesaw State University) and John Barrows (Kennesaw State University).

Our laboratory investigates transcription regulatory networks in the model extreme thermophile, Thermus thermophilus HB8. We use a reverse genetic approach, involving the iterative selection method REPSA to elucidate consensus DNA binding sequences before embarking on the identification of regulated genes. This approach utilizes extensive bioinformatics, both on the front end (e.g., PSI-BLAST and UniprotKB/PDB to identify potential test DNA-binding sequences) and on the back end (e.g., MEME Suite, KEGG, and Softberry packages to identify consensus sequences from NexGen sequencing data, place them within the T. thermophilus genome, and correlate with proposed promoters, genes, and biological functions) of our experimental work. We also mine extensive database data (e.g., NCBI GEO to identify regulated genes) deposited by other laboratories to provide correlations with our postulates and to guide additional in vitro and in vivo studies. So far, we have had considerable success identifying orphan transcriptional regulators and their networks.[1-7] Most notably, this work was primarily achieved by undergraduate students at a primarily undergraduate (teaching) institution, demonstrating the feasibility and power of incorporating bioinformatics into research experiences for undergraduates.

[1] Van Dyke MW, Beyer MD, Clay E, Hiam KJ, McMurry JL, Xie Y. Identification of preferred DNA-binding sites for the Thermus thermophilus transcriptional regulator SbtR by the combinatorial approach REPSA. PLoS One 11(7):e0159408, 7/2016. DOI: 10.1371/journal.pone.0159408 [2] Lee M, Um H, Van Dyke MW. Identification and characterization of preferred DNA-binding sites for the Thermus thermophilus transcriptional regulator FadR. PLoS One 12(9):e0184796, 9/2017. DOI: 10.1371/journal.pone.0184796 [3] Cox JS, Moncja K, Mckinnes M, Van Dyke MW. Identification and characterization of preferred DNA-binding sites for the Thermus thermophilus HB8 transcriptional regulator TTHA0973. Int J Mol Sci 20(13):3336, 7/2019. DOI: 10.3390/ijms20133336 [4] Cox JS, Van Dyke MW. General and genomic DNA-binding specificity for the Thermus thermophilus HB8 transcription factor TTHB023. Biomolecules 10(1):94, 1/2020. DOI: 10.3390/biom10010094 [5] Moncja K, Van Dyke MW. Determination and dissection of DNA-binding specificity for the Thermus thermophilus HB8 transcriptional regulator TTHB099. Int J Mol Sci 21(21):7929, 10/2020. DOI: 10.3390/ijms21217929 [6] Teague JL, Barrows JK, Baafi CA, Van Dyke MW. Discovering the DNA-binding consensus of the Thermus thermophilus HB8 transcriptional regulator TTHA1359. Int J Mol Sci 22(18):10042, 9/2021. DOI: 10.3390/ijms221810042 [7] Barrows JK, Westee AB, Parrish AY, Van Dyke MW. Identification of the preferred DNA-binding sequence and transcription regulatory network for the thermophilic zinc uptake regulator TTHA1292. J Bacteriol 204(11):e0030322, 11/2022. DOI: 10.1128/jb.00303-22

[5192] *The dilemma of choosing the best prediction model and feature engineering approach for heterogenous datasets.*

Lukasz Piorecki (Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland) and Joanna Polanska (Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland).

With so many various prediction models available nowadays, it is hard to guess which would perform the best for a given problem. Moreover, some applications require the system to be highly explainable. Thus, the main focus of the study was to create a few widely used prediction models for a nontrivial multiclass problem of diagnosing lung conditions, choose the most informative features, and compare those classification systems in terms of accuracy and explainability. The complete dataset consists of 18,962 samples with 87 numerical features, which are radiomic descriptors calculated from patients' chest X-rays pre-classified as healthy donors, nonCOVID-19 pneumonia, or COVID-19 pneumonia. The dataset was randomly split into balanced training (1,240 images per class) and hold-out (15,242 images) subsets. Three different prediction models were considered: Polytomous (Multinomial) Logistic Regression (MNR), Decision Tree (DT), and Artificial Neural Network (ANN). Using multiple random cross-validation schema (MRCV) for each prediction model, metrics and features' impact on the decisions were calculated for signature definition and further comparison. The signatures were obtained by applying the feature forward (MNRF) and feature backward (MNRB) selection algorithms in the case of MNR, the Gini-index-based scree plot analysis for DT, and SHAP scores for ANN partial models. Also, the UMAP dimension reduction technique was applied to the whole dataset to graphically determine the dataset heterogeneity. The final models were the ANNs with the signatures identified during the preliminary MRCV experiments. In addition, area-under-the-curve (AUC) and standard performance indices were calculated for each final model. It turned out that the best classification system was ANN, built on the subset of features identified by MNRF (AUC=84.19%). The second-best prediction model was constructed using the signature from MCRV experiments using MNRB (AUC=83.68%). The third top, with a slightly lower AUC value, was defined within the feature subset found by the SHAP algorithm (AUC=82.44%). The pruned Decision Tree was the worst choice of prediction system, achieving only 79.69% AUC on a hold-out set. The signatures obtained by MNRs and DT were very similar, while the one based on SHAP/ANN demonstrated higher inconsistency and required vast computational resources.

[5233] *The promise of deep learning-assisted multimodality medical image analysis*

Habib Zaidi (Geneva University Hospital).

Positron emission tomography (PET), x-ray computed tomography (CT) and magnetic resonance imaging (MRI) and their combinations (PET/CT and PET/MRI) provide powerful multimodality techniques for in vivo imaging. This talk presents the fundamental principles of multimodality imaging and reviews the major applications of artificial intelligence (AI), in particular deep learning approaches, in multimodality medical imaging. It will inform the audience about a series of advanced development recently carried out at the PET instrumentation & Neuroimaging Lab of Geneva University Hospital and other active research groups. To this end, the applications of deep learning in five generic fields of multimodality medical imaging, including imaging instrumentation design, image denoising (low-dose imaging), image reconstruction quantification and segmentation, radiation dosimetry and computer-aided diagnosis and outcome prediction are discussed. Deep learning algorithms have been widely utilized in various medical image analysis problems owing to the promising results achieved in image reconstruction, segmentation, regression, denoising (low-dose scanning) and radiomics analysis. This talk reflects the tremendous increase in interest in quantitative molecular imaging using deep learning techniques in the past decade to improve image quality and to obtain quantitatively accurate data from dedicated standalone (CT, MRI, SPECT, PET) and combined PET/CT and PET/MRI imaging systems. The deployment of AI-based methods when exposed to a different test dataset requires ensuring that the developed model has sufficient generalizability. This is an important part of quality control measures prior to implementation in the clinic. Novel deep learning techniques are revolutionizing clinical practice and are now offering unique capabilities to the clinical medical imaging community. Future opportunities and the challenges facing the adoption of deep learning approaches and their role in molecular imaging research are also addressed.

[5307] *Deep learning systems for the classification of cardiac pathologies using ECG signals*

Olga Valenzuela (University of Granada), Ignacio Rojas-Valenzuela (University of Granada), Fernando Rojas (University of Granada), Juan Carlos De la Cruz (University of Granada) and Peter Gloesekoetter (FH Muenster University of Applied Sciences).

n this paper, several deep learning models are analyzed for the construction of the automated helping system to ECG classification. The methodology presented in this article begins with a study of the different alternatives for performing the discrete wavelet transform-based scalogram for an ECG. Then, several Deep Learning architectures are analysed. Due

to the large number of architectures in the literature, seven have been selected as they have a high degree of acceptance in the scientific community. The influence of the number of epochs used for training will also be analysed.

In addition to the development of a classifier able to accurately solve the multi-class problem of, given an ECG signal, deciding which pathology the subject is suffering from (main interest for a medical expert), we also want to rigorously analyze, through the use of a statistical tool (ANOVA), the impact the main functional blocks of our system have on its behaviour.

As a novel result of this article, different homogeneous groups of deep learning systems are analysed (from a statistical point of view, they have the same impact on the accuracy of the system). As can be seen in the results, there are four homogeneous groups, with the group with the lowest accuracy index obtaining an average value of 76,48% in the classification and the group with the best results, with an average accuracy of 83,83%.

[5387] *A Platform for the Study of Drug Interactions and Adverse Effects Prediction*

Diogo Mendes (University of Porto Faculty of Engineering) and Rui Camacho (University of Porto Faculty of Engineering).

This article reports on the development of a Web platform for the study of Adverse Drug Events (ADEs). The platform is able to import ADE episodes from oficial Web sites, like OpenFDA, analyze the chemistry of the drugs involved, together with patient data, and produce a potential explanation based on the drugs interactions. Each study uses chemical knowledge to enrich the information on the molecules involved in the episodes. Data Mining is then used to construct models that can help in the explanation of the ADE occurrence and to predict future events. This paper reports on the Web portal developed and the Data Mining experiments conducted to evaluate the quality, and potential explanations of the forecasted adverse reactions, using real reports of drug administration and the subsequent adverse events. The results showed that it was possible to predict the outcomes of ADEs based on the structure of the molecules of the drugs involved and the data collected from real reports of drug administration up to an accuracy of 79%, while also predicting, with high accuracy, the seriousness of events where the outcome is the death of the patient (with a precision of 98.9%). The platform provides a less expensive and more accurate way of predicting adverse drug reactions compared to traditional methods. This study highlights the importance of understanding drug interactions at a molecular level and the usefulness of using Data Mining techniques in predicting ADEs.

[5421] *The effectiveness of quarantine in viral and bacterial epidemics: new evidence provided by the Covid-19 pandemic*

Andreu Martínez-Hernández (Hospital General de Castelló, 12004 Castelló de la Plana, Spain) and Vicente Martínez (Institut de Matemàtiques i Aplicacions de Castelló, Departament de Matemàtiques, Universitat Jaume I).

The effectiveness of confining the population has been observed for centuries. However, this effectiveness has now been demonstrated with data during the COVID-19 pandemic, for which an enormous amount of data and studies are available. In this sense, this paper identifies the determination of the number of people susceptible to contracting the disease, which is present in many epidemic transmission models, as the fundamental variable for understanding the dynamics of infections. We primarily consider the SIRD model, but the data are also contrasted with models and techniques used in pandemic analysis. In addition, the facts and conditions of the COVID-19 pandemic are compared with others that occurred historically.

[5473] *Digital Breast Tomosynthesis reconstruction techniques in healthcare systems: A review*

Imane Samiry (Hassan II university), Ilhame Ait Lbachir (University Hassan II), Imane Daoudi (University Hassan II), Saida Tallal (University Hassan II) and Sayouti Adil (University Hassan II).

Digital Breast Tomosynthesis images are widely used to increase breast cancer detection and reduce recall rates in healthcare systems for breast cancer detection. In the field of medical imaging, computer-aided diagnosis (CAD) systems are used to analyze this type of images. Generally, in order to achieve an early detection of breast cancer, these CAD systems start with the reconstruction part of the image, the pre-processing step and then the segmentation and classification. However, the post-acquisition techniques of Digital Breast Tomosynthesis can impact the detection and diagnosis of breast cancer, and bias the final decision in computer-aided detection and diagnosis systems. Mainly, the recostruction phase in computer aided detection systems, that helps prepare the DBT for further analysis, such as segmentation and classification of abnormalities. In this paper, we present a survey of different techniques for digital

breast tomosynthesis reconstruction, that we compared theoretically in terms of advantages and drawbacks, particularly for healthcare systems dedicated to breast cancer detection.

[5477] *Smart Wearables Data Collection and Analysis for Medical Applications: A Preliminary Approach for Functional Reach Test*

João Duarte (Polytechnic Institute of Leiria), Luis Francisco (Polytechnic Institute of Leiria), Ivan Miguel Pires (Instituto de Telecomunicações) and Paulo Coelho (Polytechnic Institute of Leiria).

The Functional Reach Test (FRT) is a commonly used clinical tool to evaluate the dynamic balance and fall risk in older adults and individuals with specific neurological conditions. Several studies have highlighted the importance of using FRT as a reliable and valid measure for assessing functional balance and fall risk in diverse populations. Additionally, FRT is sensitive to changes in balance function over time and can provide critical information for designing rehabilitation programs to improve balance and reduce the risk of falls. The FRT has also been used as a screening tool for identifying individuals who may benefit from further assessment or intervention. Thus, the FRT is a valuable clinical instrument for assessing functional balance and fall risk and should be incorporated into routine clinical practice. This paper intends to describe the preliminary results and future directions for implementing the FRT with various sensors gathered from smartphones or smart wearables to provide valuable indicators to aid professional healthcare practitioners in evaluating and following up on the elderly, but possibly extending to other age groups.

[5481] *A Machine Learning approach to predict brain abnormalities in preterm infants using clinical data*

Arantxa Ortega Leon (Universidad de Cádiz), Roa'A Khaled (Universidad de Cádiz), María Inmaculada Rodríguez García (Universidad de Cádiz), Daniel Urda (University of Burgos) and Ignacio Turias (Universidad de Cádiz).

Preterm infants are prone to several neurodevelopmental impairments (NDI). Early and accurate diagnosis could cooperate in the treatment of their clinical manifestations. Clinical data from a cohort of preterm infants from the Hospital Puerta del Mar, Cádiz, Spain was used in this work to perform a classification task to predict abnormal MRI findings using clinical data. In this paper, we describe an approach using Machine Learning models to predict abnormal MRI findings using clinical data collected at the NICU. Results in this analysis indicate that the best model able to predict abnormal MRI findings was K-nearest Neighbor, with a recall of 0.80. This study represents an initial step towards developing a practical and reliable tool for predicting abnormal MRI findings in preterm infants using readily available clinical data from preterm infants.

[5581] *Analysis of the organization of the structure of human genes depending on their tissue specificity*

Sergey Slyusarev (Southern Federal University) and Olga Lyangasova (Southern Federal University).

Human genes can be divided into ubiquitously expressed genes, and genes that are expressed in only one type of cell or tissue. A small number of human genes have strict cellular or tissue monospecificity, however, understanding the organization of the gene structure during mono- and poly-tissue expression is an important task of modern genetics. This work aims to find structural differences between tissue-specific genes and genes that are ubiquitously expressed (housekeeping genes). To achieve this goal, 40 highly tissue-specific genes and 40 housekeeping genes were selected. The work was carried out using the NCBI genome browser, Ensemble genome browser, as well as gene expression databases (GTEx, Illumina, BioGPS). The determination of the phases of introns was carried out using the author's program.
The number of introns with different phases did not differ in the compared groups. The presence of CpG islands within the boundaries of the studied genes, in the 5'-UTR region, was studied. All selected housekeeping genes were characterized by the presence of a GC island in the 5' UTR region. For tissue-specific genes, the presence of GC islands within the 5'-UTR region of the gene was not detected.

[5608] *The dark side of NCBI: Annotation artifacts across the RefSeq database*

Martin Bartas (Department of Biology and Ecology, Faculty of Science, University of Ostrava), Jiří Červeň (Department of Biology and Ecology, Faculty of Science, University of Ostrava), Adriana Volná (Department of Physics, Faculty of Science, University of Ostrava, Ostrava, Czech Republic) and Petr Pečinka (Department of Biology and Ecology, Faculty of Science, University of Ostrava).

Current bioinformatical approaches largely rely on data deposited in various databases, among which the National Center for Biotechnology Information (NCBI) occupies a privileged place. Thousands of scientific articles present data based on sequences obtained from NCBI each year. The Reference Sequence (RefSeq) database, according to the original statement made by the National Institutes of Health (NIH) is the 'comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic, transcript, and protein'. Therefore, most scientists naturally consider these data to be reliable and extensively use them in various types of bioinformatic analyses. In this study, we would like to show some recent data pointing to the fact that the RefSeq database is full of annotation artifacts, comprising many examples across the whole tree of life. Specifically, we have focused e.g. on the protein sequences of well-known tumor suppressor p53 that are in some cases artificially extended by 5´-UTRs, or artificial chalcone synthase (CHS) domain multiplication in some important plant species. Our study aims to raise awareness about annotation errors concerning the NCBI RefSeq database and suggest a workflow on how to deal with this issue.

[5616] *Predicting Cancer Stage from Circulating microRNA: A Comparative Analysis of Machine Learning Algorithms*

Sören Richard Stahlschmidt (University of Skövde), Benjamin Ulfenborg (University of Skövde) and Jane Synnergren (University of Skövde).

In recent years, serum-based tests for early detection and detection of tissue of origin are being developed. Circulating microRNA has been shown to be a potential source of diagnostic information that can be collected non-invasively. In this study, we investigate circulating microRNAs as predictors of cancer stage. Specifically, we predict whether a sample stems from a patient with early stage (0-II) or late stage can- cer (III-IV). We trained four machine learning algorithms on a data set of cancers from twelve different primary sites. The results showed that cancer stage can be predicted from circulating microRNA with a sensi- tivity of 70.61% and positive predictive value of 51.40%. Furthermore, we compared the best pan-cancer model with models specialized on indi- vidual cancers and found no statistically significant difference. Finally, in the best performing pan-cancer model 533 microRNAs were significant. Comparing the five most relevant circulating microRNAs found with the current literature showed clear associations to various cancers. In con- clusion, the study showed the potential of circulating microRNA and machine learning algorithms to predict cancer stage and thus suggests that further research into its potential as a non-invasive clinical test is warranted

---

[5617] *Radar Sensing in Healthcare: Challenges and Achievements in Human Activity Classification & Vital Signs Monitoring*

Francesco Fioranelli (TU Delft), Ronny Guendel (TU Delft), Nicolas Kruse (TU Delft) and Alexander Yarovoy (TU Delft).

Driven by its contactless sensing capabilities and the lack of optical images being recorded, radar technology has been recently investigated in the context of human healthcare. This includes a broad range of applications, such as human activity classification, fall detection, gait and mobility analysis, and monitoring of vital signs such as respiration and heartbeat. In this paper, a review of notable achievements in these areas and open research challenges is provided, showing the potential of radar sensing for human healthcare and assisted living.

[5730] *The effect of biofeedback on learning the wheelie position on manual wheelchair*

Antonio Pinti (LARSH DeVISU UPHF Valenciennes).

The aim of this study was to investigate the impact of biofeedback (BFB) on manual wheelchair learning. The researchers conducted training sessions with two groups of participants, one using BFB and the other group without it (NBFB). The hypothesis was that BFB would reduce the learning time and help participants to achieve balance positions more quickly. The study enrolled 24 participants aged 24± 6 years old; they were divided into two groups of 12 subjects each (BFB and NBFB). The researchers also collected additional information about the participants, such as the sport they practiced, for future investigations. The data was collected using a non-contact electronic angular system placed directly on the

wheelchair, measuring spatiotemporal parameters such as the angle between the wheelchair and the ground and the time at which this angle is reached. The results which are statistically significant ($p < 0.05$) were only obtained between early falling, learning time and number of trials. The study found that BFB did not seem to accelerate the learning time for the wheelie skill on manual wheelchair (BFB group). However, the BFB method could potentially reduce the number of trials using the manual wheelchair under (NBFB). In conclusion, the study showed that biofeedback may not necessarily accelerate the learning time for the wheelie skill on manual wheelchair but can help individuals to maintain balance positions with fewer trials. Further studies are required to confirm these results, as they only involved a small sample size. This study highlights the potential for using biofeedback as an effective tool for wheelchair training and could improve the quality of life of individuals with mobility impairments.

[5739] *Ethical dilemmas, mental health, artificial intelligence and LLM based chatbots*

Johana Cabrera (University of Santiago Chile), Soledad Loyola (University of Santiago de Chile), Irene Magaña (University of Santiago de Chile) and Rodrigo Rojas (University of Santiago de Chile).

With the passage of time, many tasks that were once considered human tasks have gradually been replaced by machines. This phenomenon has been widely observed since the Industrial Revolution [1].

Currently, we are said to be experiencing the Fourth Industrial Revolution, which is marked by a series of technological, digital, physical, and even biological convergences, including the development of the internet, automation, robotics, blockchain, cloud computing, 3D printing, and artificial intelligence [2].

The technology that, in the last period, has caused a high controversy, is Artificial Intelligence (AI), which has been defined by John McCarthy, as "the combination of science and engineering to make intelligent devices for human welfare"[3].

The current debate is generated around AI, which has been sparked by one of its derivatives, Chatbots, based on large language models (LLM). These chatbots are based on a language model that allows the deep and accelerated learning of such technologies [4]. Among these, ChatGPT-3 and ChatGPT-4 [5], which have caused the greatest impact so far, can be mentioned because of their great capacity both in function and interaction with human beings. Such is the evolution of this technology that Bill Gates has called "the most important technological advance since the graphical user interface [6].

It is important to consider that this scene takes place in a delicate context for humanity, since, in recent years, the world is going through a series of economic, health, and sociopolitical problems at a national and international level, which will also have a reciprocal impact, affecting the health and mental health of people, as well as enhancing the progress of the scientific-technological devices[2, 9, 10].

Even though the technology is neutral, its use involves both potential benefits and risks. This reality shall converge different actors in society to make the best use of it.

Due to the above, immediate actions are required such as ethical reflections, and scientific development, that shed light on the developments, uses, and regulations, around what Gates states as the “new era, the age of Artificial Intelligence has begun" [6], which we know will continue to create changes in our society.

Some experts state that in the interim these chatbots, will have an impact on 19% of the work done by humans [7], continue to spread the bias of their creators, and make advancements in their persuasive interaction in the political and marketing fields, impacting directly on the thoughts, emotions, and behaviors of the population [8].

In this sense, Bill Gates [6] announced two important fields of development: global health and education. This implies an investment in the formation of advanced human capital and in improving access for the population to these types of tools, which would constitute part of the major actions to be developed.

In this context, the field of mental health is being affected by these technologies in various axes of advice and professional functions. These are areas such as diagnosis, intervention design, monitoring, treatment, accompaniment, interaction, referral decisions, and the development of support material, which is to be supposed be enhanced and improved by these chatbots. An example of this would be that the tool has achieved a 19% increase in conversational empathy when people use it to generate texts [13].

The aim of this paper is to reflect, on the proposals of the bioethical principles (beneficence, nonmaleficence, justice, and autonomy) [19] and ethical dilemmas involved in the use of the LLM chatbots in the professional practice of mental health.

For this purpose, a thematic analysis and systematization of the information produced in documents both in the scientific field and in some mass media communications will be carried out. The results aim to provide relevant information through the production of protocols and recommendations that will serve at an interdisciplinary level, for inputs to regulation, use, research, and training.

[5863] *Uterine Cervix and Corpus Cancers Characterization through Gene Expression Analysis Using the KnowSeq Tool*

Lucía Almorox (University of Granada), Luis Javier Herrera (University of Granada), Francisco Ortuño (University of Granada) and Ignacio Rojas (University of Granada).

The characterization of cancer through gene expression quantification data analysis is a powerful and widely used approach in cancer research. This paper describes two experiments that demonstrate its potential in identifying differentially expressed genes (DEGs) and accurately predicting cancer subtypes. To achieve this, RNA-seq data was obtained from TCGA database and subsequently preprocessed and analyzed using the KnowSeq package from Bioconductor. In the first experiment, the study focuses on identifying DEGs in healthy, cervical cancerous, and uterine corpus cancerous tissues. The kNN classifier was employed to evaluate the utility of these genes in predicting a sample belonging to one of these three classes. A gene signature consisting of only three genes produced remarkable results on a 5-fold cross-validation assessment process, with overall test accuracy and F1 values of 99.33% and 96.73%, respectively. The paper provides ontological enrichment, associated diseases, and pathways of the gene signature to shed light on the molecular mechanisms involved in both cancers. The second experiment extends the work by classifying cervical cancer samples into their two most common histological types: adenocarcinoma and squamous cell carcinoma. By using a single gene, the study was able to achieve 100% of test accuracy in a 5-fold cross-validation process. Additionally, the classification of an adenosquamous sample into one of these two categories based on the number of genes used was also examined. Overall, these experiments demonstrate the potential of these techniques to improve cancer diagnosis and treatment. Moreover, the study provides valuable insights into the underlying molecular mechanisms of cervix and uterine corpus cancers, laying the groundwork for further research in this field.

[5952] *Unsupervised investigation of information captured in pathway activity score in scRNA-Seq analysis*

Kamila Szumala (Silesian University of Technology), Joanna Polanska (Silesian University of Technology) and Joanna Zyla (Silesian University of Technology).

With the introduction of single cell RNA sequencing, research on cell, tissue and disease heterogeneity has a new boost. Transforming gene levels to explainable pathways via single-sample enrichment algorithms is a leading analysis step in understanding cell heterogeneity. In this study, eight different single-sample methods were investigated and accompanied by gene level outcomes as reference. For all, their ability to cell separation and clustering accuracy was tested. For this purpose, six scRNA-Seq datasets with labelled cells and their various numbers were collected. PLAGE method shows the best cell separation with statistically significant differences to gene level and to six other tested methods. The clustering accuracy analysis also indicates that PLAGE is the leading technique in single-sample enrichment methods in scRNA-Seq. Here the worst performance was observed to JASMINE algorithm which, in contrary to PLAGE, was designed to analyse the scRNA-Seq data. Moreover, Louvain clustering shows the best results regarding cell division regardless of the tested single-sample method. Finally, the results of clustering given by PLAGE reveal T cell subtypes initially not labelled, showing the great potential of this algorithm in heterogeneity investigation.

[6035] *MetaLLM: Residue-wise Metal ion Prediction Using Deep Transformer Model*

Fairuz Shadmani Shishir (University of Kansas), Bishnu Sarker (Meharry Medical College), Farzana Rahman (Kingston University London) and Sumaiya Shomaji (University of Kansas).

Proteins bind to metals such as copper, zinc, magnesium, etc., serving various purposes such as importing, exporting, or transporting metal in other parts of the cell as ligands and maintaining stable protein structure to function properly. A metal binding site indicates the single amino acid position where a protein binds a metal ion. Manually identifying metal binding sites is expensive, laborious, and time-consuming. A tiny fraction of the millions of proteins in UniProtKB – the most comprehensive protein database – are annotated with metal binding sites, leaving many millions of proteins waiting for metal binding site annotation. Developing a computational pipeline is thus essential to keep pace with the growing number of proteins. A significant shortcoming of the existing computational methods is the consideration of the long-term dependency of the residues. Other weaknesses include low accuracy, absence of positional information, hand-engineered features, and a pre-determined set of residues and metal ions. In this paper, we propose MetaLLM, a metal binding site prediction technique, by leveraging the recent progress in self-supervised attention-based (e.g. Transformer) large language models (LLMs) and a considerable amount of protein sequences publicly available. LLMs are capable of

modelling long residual dependency in a sequence. The proposed MetaLLM uses a transformer pre-trained on an extensive database of protein sequences and later fine-tuned on metal-binding proteins for multi-label metal ions prediction. A stratified 10-fold cross-validation shows more than 90% precision for the most prevalent metal ions. Moreover, the comparative performance analysis confirms the superiority of the proposed MetaLLM over classical machine learning techniques.

[6114] *Complex Network and Artificial Intelligence combined approach to investigate Autism spectrum disorder through gene expression data*

Antonio Lacalamita (Università degli studi di Bari Aldo Moro), Alfonso Monaco (Università degli studi di Bari Aldo Moro), Nicola Amoroso (Università degli studi di Bari Aldo Moro), Loredana Bellantuono (Università degli studi di Bari Aldo Moro), Alessandro Fania (Università degli studi di Bari Aldo Moro), Ester Pantaleo (Università degli studi di Bari Aldo Moro), Sabina Tangaro (Università degli studi di Bari Aldo Moro) and Roberto Bellotti (Università degli studi di Bari Aldo Moro).

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that has a complex etiology in which only 20% of cases are explained by known genomic mutations. In recent years, transcriptome analysis between the autistic and normal brains has identified the association between gene expressions and disorder. In our study we analyzed a publicly available gene expression dataset composed of 94 samples (51 healthy subjects and 43 autistic subjects), in order to find the genetic communities that are most related to the ASD phenotype. Through a data-driven approach based on complex networks, we have grouped thousands of genes into small stable communities by exploiting the spin-glass algorithm inspired by the model of Nobel Prize winner Giorgio Parisi. We then applied the Boruta algorithm, a wrapper method, to select the most discriminating genes within each of the 29 found communities. Finally, with the selected genes we fed a Machine Learning Random Forest (RF) algorithm within a cross-validation framework. RF provides a classification accuracy for the 29 found communities ranging from 70.86% to 77.61%. Our results suggest that the combined use of brain gene expression and artificial intelligence may help find new ASD-related biomarkers. Further studies are needed to biologically and statistically validate these gene communities.

[6258] *Meta-analysis of gene activity (MAGA) contributions and correlation with gene expression, through GAGAM.*

Lorenzo Martini (Politecnico di Torino, Control and Computer Engineering Department), Roberta Bardini (Politecnico di Torino, Control and Computer Engineering Department), Alessandro Savino (Politecnico di Torino, Control and Computer Engineering Department) and Stefano Di Carlo (Politecnico di Torino, Control and Computer Engineering Department).

It is well-known how sequencing technologies propelled cel- lular biology research in the latest years, giving an incredible insight into the basic mechanisms of cells. Single-cell RNA sequencing is at the front in this field, with Single-cell ATAC sequencing supporting it and becoming more popular. In this regard, multi-modal technologies play a crucial role, allowing the possibility to perform the mentioned sequenc- ing modalities simultaneously on the same cells. Yet, there still needs to be a clear and dedicated way to analyze this multi-modal data. One of the current methods is to calculate the Gene Activity Matrix, which summarizes the accessibility of the genes at the genomic level, to have a more direct link with the transcriptomic data. However, this concept is not well-defined, and it is unclear how various accessible regions impact the expression of the genes. Therefore, this work presents a meta-analysis of the Gene Activity matrix based on the Genomic-Annotated Gene Ac- tivity Matrix model, aiming to investigate the different influences of its contributions on the activity and their correlation with the expression. This allows having a better grasp on how the different functional regions of the genome affect not only the activity but also the expression of the genes.

[6449] *Transparent Machine Learning Algrithms for Explainable AI on Motor fMRI Data*

Jose Diogo Marques dos Santos (University of Porto), David Machado (University of Porto) and Manuel Fortunato (University of Porto).

With the emergence of explainable artificial intelligence (xAI), two main approaches for tackling model explainability have been put forward. Firstly, the use of inherently simple and transparent models that with easily understandable inner-workings (interpretability) and can readily provide useful knowledge about the model's decision making process

(explainability). The second approach is the development of interpretation and explanation algorithms that may shed light upon black-box models. This is particularly interesting to apply on fMRI data as either approach can provide pertinent information about the brain's underlying processes. This study aims to explore the capability of transparent machine learning algorithms to correctly classify motor fMRI data, if more complex models inherently lead to a better prediction of the motor stimulus, and the capability of the Integrated Gradients method to explain a fully connected artificial neural network (FCANN) used to model motor fMRI data. The transparent machine learning models tested are Linear Regression, Logistic Regression, Naive Bayes, K-Neighbors, Support Vector Machine, and Decision Tree, while the Integrated Gradients method is tested on a FCANN with 3 hidden layers. It is concluded that the transparent models may accurately classify the motor fMRI data, with accuracies ranging from 66.75 \% to 85.0 \%. The best transparent model, multinomial logistic regression, outperformed the most complex model, FCANN. Lastly, it is possible to extract pertinent information about the underlying brain processes via the Integrated Gradients method applied to the FCANN by analyzing the spatial expression of the most relevant Independent Components for the FCANN's decisions.

[6459] *Predicting the Risk of Recurrence and Prognosis in Patients with Hepatocellular Carcinoma*

Chi-Chang Chang (chungshan medical university).

Background: Recurrence is one of the major clinical problems for survivors of hepatocellular carcinoma (HCC). A comprehensive analysis of evidence-based medicine is currently necessary to prevent the recurrence of HCC.

Methods: A retrospective analysis of 5,665 records from five cancer registries was conducted, and 14 candidate risk factors were selected based on literature and clinical experts, which are age at diagnosis, sex, tumor size, tumor number, Barcelona Clinic Liver Cancer (BCLC) stage, curative treatment, body mass index (BMI), smoking behavior, betel nut chewing behavior, alcohol use, alpha-fetoprotein (AFP), total bilirubin, hepatitis B virus (HBV), hepatitis C virus (HCV). Statistical and machine learning classifiers were used to select the important features. In addition, five predictive models (CART (Classification And Regression Tree), C4.5 (C4.5 Decision tree), LDA (Linear Discriminant Analysis), RF (Random Forest) and LGR (Logistic Regression)) were used to analyze different recurrence types and treatment modalities. We evaluated the predictive models based on their sensitivity, specificity, accuracy, F-measure score, Kappa, Matthews correlation coefficient, and area under the curve (AUC).

Results: The three most important risk factors for overall recurrence were tumor number, BCLC stage and curative treatment. For local regional recurrence, surgical procedure, tumor number and BCLC stage were the three most common risk factors; relative to local regional recurrence, tumor size, AFP and tumor number were the three most common risk factors of distant recurrence. According to the results of the feature selection for different treatment modalities, the most common risk factors for surgical resection were pathologic stage, tumor size, and tumor number. For radiofrequency ablation (RFA), the three main risks were tumor number, tumor size, and BCLC stage; for transcatheter arterial chemoembolization (TACE), the three main risks were HCV, tumor number, and betel nut chewing behavior. In all prediction modes, C4.5 had the highest AUC value, between 0.7573 and 0.9675. Discussions and Conclusion: The results of this study showed that tumor number was the common risk factor. The risk factors for overall recurrence, tumor number, AFP, and tumor size were consistent with previous findings, but sex and BCLC Stage were different. The risk factors for the different types of recurrence, tumor size and AFP were consistent with previous findings, while tumor number and BCLC stage were not. The risk factors for the surgical resection treatment were consistent with previous findings, including tumor size, tumor number, AFP, grade and surgical margins. The RFA results showed that tumor number, tumor size, and AFP were consistent with previous findings; however, age and sex were different. In TACE, tumor number and BCLC stage were consistent with previous findings; however, tumor size and AFP were not. From an empirical medical perspective, this study uses machine learning classifiers and statistical methods to analyze the Taiwan cancer registry database. Furthermore, the selected risk factors can be used as predictive models for clinical decision-making. This will allow physicians to identify disease risks and provide the best monitoring and tracking of clinical adjustment.

[6505] *Inter-helical residue contact prediction in α-helical Transmembrane proteins using structural features*

Aman Sawhney (University of Delaware), Jiefu Li (University of Delaware) and Li Liao (University of Delaware).

Residue contact maps offer a 2-d, reduced representation of 3-d protein structures and constitute a structural constraint and scaf- fold in structural modeling. Precise residue contact maps are not only helpful as an intermediate step towards generating effective 3-d protein models, but also useful in their own right in identifying binding sites and hence providing

insights about a protein's functions. Indeed, many computational methods have been developed to predict residue contacts using a variety of features based on sequence, physio-chemical properties, and co-evolutionary information. In this work, we set to explore the use of structural information for predicting inter-helical residue contact in transmembrane proteins. Specifically, we extract structural information from a neighborhood around a residue pair of interest and train a clas- sifier to determine whether the residue pair is a contact point or not. To make the task practical, we avoid using the 3-d coordinates directly, in- stead we extract features such as relative distances and angles. Further, we exclude any structural information of the residue pair of interest from the input feature set in training and testing of the classifier. We compare our method to a state-of-the-art method that uses non-structural infor- mation on a benchmark data set. The results from experiments on held out datasets show that the our method achieves above 90% precision for top L/2 and L inter-helical contacts, significantly outperforming the state-of-the-art method and may serve as an upper bound on the per- formance when using non-structural information. Further, we evaluate the robustness of our method by injecting Gaussian normal noise into PDB coordinates and hence into our derived features. We find that our model's performance is robust to high noise levels.

[6563] *Preliminary Results of Using the Tangram Meta-Heuristic for Virtual Screening in Drug Discovery*

N.C. Cruz (University of Granada), S. Puertas-Martin (University of Almería), J.L. Redondo (University of Almería) and P.M. Ortigosa (University of Almería).

Virtual screening methods focus on looking for the most similar compounds to a given one in massive compound databases. They allow accelerating and reducing costs when designing new drugs. In this field, shape similarity is one of the most used metrics. However, its computation requires finding the optimal comparison position between the query and every studied or candidate molecule. The positioning procedure is addressed as an optimization problem. As compound databases are massive, this problem must be solved as fast as possible, which generally implies using local optimizers. Recently, a global evolutionary method has been proposed with promising results, OptiPharm. Unfortunately, it has multiple parameters, it is complex due to its evolutionary background, and its computational budget is high. This work explores the possibility of using the more recent optimizer Tangram instead. It shares the origin but has been designed to be simpler to tune and to require fewer function evaluations. According to the preliminary results, in 57.5% of the studied cases, Tangram finds the same or equivalent compounds for a given query with a tenth of the computational budget.

[6568] *Investigation of inclusion for localised characteristics from medical imaging datasets genotype-phenotype associations*

Gabrielle Dagasso (University of Calgary), Matthias Wilms (University of Calgary) and Nils Forkert (University of Calgary).

Advances in bioinformatics have led to an improved understanding of geno-type-phenotype associations in the human body. However, the usage of neu-roimaging data as a phenotype is still mostly limited to using derived meas-urements or dimensionality reduction methods rather than using the full im-ages due to high dimensionality of imaging datasets. Therefore, important detailed information in the images may be missed in such studies. In this work, we propose a flexible methodology based on morphometrics for dimensionality reduction that can make use of localized as well as more global imaging information. The morphometrics approach was applied to data from patients with attention-deficit hyperactivity disorder, a hereditary disorder characterized by known structural changes in the brain with different dis- tance thresholds. The distances were compared with the results of a standard multivariate genome-wide association study. The results showed that the proposed dimensionality reduction approach is able to retain local imaging information within individual features. In summary, using this approach enables one to efficiently and robustly include whole medical imaging data into multivariate genome-wide association studies without the need for pre-specification of brain region measurements, offering a novel approach to incorporating full medical imaging data as phenotypes.

[6799] *A guide and mini-review on the performance evaluation metrics in binary segmentation of magnetic resonance images*

Ayca Kirimtat (University of Hradec Kralove) and Ondrej Krejcar (University of Hradec Kralove).

Eight previously proposed segmentation evaluation metrics for brain magnetic resonance images (MRI), which are sensitivity (SE), specificity (SP), false-positive rate (FPR), false-negative rate (FNR), positive predicted value (PPV), accuracy (ACC), Jaccard index (JAC) and dice score (DSC) are presented and discussed in this paper. These evaluation metrics could be classified in two groups namely as pixel-wise metrics and area-wise metrics. We, also, distill the most prominent previously published papers on brain MRI segmentation eval-uation metrics between 2021 and 2023 in a detailed literature matrix. The identi-fication of illness or tumor areas using brain MRI image segmentation is a large area of research. However, there is no single segmentation evaluation metric when evaluating the results of brain MRI segmentation in the current literature. Also, the pixel-wise metrics should be supported with the area-wise metrics such as DSC while evaluating the image segmentation results and each metric should be compared with other metrics for better evaluation.

[6833] *Data Augmentation Techniques for Improving Biomedical Name Entity Recognition in Low-Resource Settings*

Yiling Cao (FUJITSU R&D CENTER CO.,LTD.), Zhongguang Zheng (Fujitsu R&D Center Co., Ltd.) and Lu Fang (Fujitsu R&D Center Co., Ltd.).

In this paper, we focus on the task of data augmentation for biomedical domain Spanish named entity recognition in low-resource scenarios. Due to the complexity nature of biomedical terms in Spanish and the lack of labeled samples, this task remains challenging. To address these issues, we introduce a new data augmentation method based on a two-step LSTM-based language model, which contains a training step and a generation step. Specifically, in the training step, we successively do the synonym replacement strategy to obtain augmented data and train the language model with the original and augmented samples. In the generation step, we first generate the synthetic data with a trained language model, and then filter out the low-quality samples via rule-based strategies and a model-based method. We evaluate our method on the biomedical domain Spanish named entity recognition dataset and the results show that our approach outperforms the baseline model.

[6924] *Degree-normalization improves random-walk-based embedding accuracy in PPI graphs*

Luca Cappelletti (Dipartimento di Informatica - Università degli Studi di Milano), Stefano Taverni (Dipartimento di Informatica - Università degli Studi di Milano), Tommaso Fontana (Dipartimento di Informatica - Università degli Studi di Milano), Marcin P. Joachimiak (Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA), Justin Reese (Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA), Peter Robinson (The Jackson Laboratory for Genomic Medicine, Farmington, USA), Elena Casiraghi (Dipartimento di Informatica - Università degli Studi di Milano) and Giorgio Valentini (Dipartimento di Informatica, Università degli Studi di Milano).

Among the many proposed solutions in graph embedding, traditional random walk-based embedding methods have shown their promise in several fields. However, when the graph contains high-degree nodes, random walks often neglect low- or middle-degree nodes and tend to prefer stepping through high-degree ones instead. This results in random-walk samples providing a very accurate topological representation of neighbourhoods surrounding high-degree nodes, which contrasts with a coarse-grained representation of neighbourhoods surrounding middle and low-degree nodes. This in turn affects the performance of the subsequent predictive models, which tend to overfit high-degree nodes and/or edges having high-degree nodes as one of the vertices. We propose a solution to this problem, which relies on a degree normalization approach. Experiments with popular RW-based embedding methods ap- plied to edge prediction problems involving eight protein-protein interaction (PPI) graphs from the STRING database show the effectiveness of the proposed approach: degree normalization not only improves predictions but also provides more stable results, suggesting that our proposal has a regularization effect leading to a more robust convergence.

[6932] *Recent Advances in Discovery of New Tyrosine Ki-nase Inhibitors Using Computational Methods*

Vesna Rastija (Faculty of Agrobiotechnical Sciences Osijek) and Maja Molnar (Faculty of Food Technology Osijek).

The development of new inhibitors of tyrosine kinase and the study of their mode of action is a necessary part of the discovery of a new cancer therapy. However, drug discovery and development is a very complex, time-consuming, and expensive process. Rational drug design methods, which comprise computational techniques, minimize that time and

costs. In this article, we reviewed recent advances in the study of tyrosine kinase inhibitors performed by computational techniques, such as quantitative structure-activity relationship (QSAR), and molecular docking.

[6973] *Identification of InhA-inhibitors interaction fingerprints that affect residence time*

Magdalena Ługowska (Silesian University of Technology) and Marcin Pacholczyk (Silesian University of Technology).

Drug development is a complex process that remains subject to risks and uncertainties. In its early days, much emphasis was placed on the equilibrium binding affinity of a drug to a particular target, which is described by the equilibrium dissociation constant (Kd). However, there are a large number of drugs that exhibit non-equilibrium binding properties. For this reason, optimization of other kinetic parameters such as dissociation constants (kof f) and association constants (kon) is becoming increasingly important to improve accuracy in measuring in vivo efficacy. To achieve this, the concept of residence time between drug and target (τ) was developed to account for the continuous elimination of the drug, the absence of equilibrium conditions, and the conformational dynamics of the target molecules. Residence time has been shown to be a better estimate of drug lifetime potency than equilibrium binding affinity and is recognized as a key parameter in drug development. However, because residence time is only one measure of drug potency, it provides only a limited picture of binding kinetics and affinity. A machine-learning algorithm was proposed to identify molecular features affecting protein-ligand binding kinetics for a set of similar compounds. Molecular dynamics simulations of τ RAMD results were used as model input. The study confirmed that τ RAMD provides information about the characteristics of the dissociation pathway since the obtained dissociation trajectories can be used to identify the interactions that occur and the conformational changes of the system at subsequent time points. The proposed algorithm made it possible to obtain information on protein-ligand contacts that are specific to their residence times.

[7008] *A Meta-Graph for the Construction of an RNA-centered Knowledge Graph*

Emanuele Cavalleri (Dipartimento di Informatica - Università degli Studi di Milano), Sara Bonfitto (Dipartimento di Informatica - Università degli Studi di Milano), Alberto Cabri (Dipartimento di Informatica - Università degli Studi di Milano), Jessica Gliozzo (Dipartimento di Informatica - Università degli Studi di Milano), Paolo Perlasca (Dipartimento di Informatica - Università degli Studi di Milano), Mauricio Soto-Gomez (Dipartimento di Informatica - Università degli Studi di Milano), Gabriella Trucco (Dipartimento di Informatica - Università degli Studi di Milano), Elena Casiraghi (Dipartimento di Informatica - Università degli Studi di Milano), Giorgio Valentini (Dipartimento di Informatica - Università degli Studi di Milano) and Marco Mesiti (Dipartimento di Informatica - Università degli Studi di Milano).

The COVID-19 pandemic highlighted the importance of RNA-based technologies for the development of new vaccines. Besides vaccines, a world of RNA-based drugs, including small non-coding RNA, could open new avenues for the development of novel therapies covering the full spectrum of the main human diseases. In the context of the “National Center for Gene Therapy and Drugs based on RNA Technology” funded by the Italian PNRR and the NextGenerationEU program, our lab will contribute to the construction of a Knowledge Graph (KG) for RNA-drug analysis and the development of innovative algorithms to support RNA-drug discovery. In this paper, we describe the initial steps for the identification of public data sources from which information about different kinds of non-coding RNA sequences (and their relationships with other molecules) can be collected and used for feeding the KG. An in-depth analysis of the characteristics of these sources is provided, along with a meta-graph we developed to guide the RNA-KG construction by exploiting and integrating biomedical ontologies and relevant data from public databases.

[7038] *Physiological polyphosphate: a new molecular paradigm in biomedical applications for human therapy*

Prof. Dr. Werner E. G. Müller (University Medical Center of the Johannes Gutenberg University Mainz) and Prof. Dr. Xiaohong Wang (University Medical Center of the Johannes Gutenberg University Mainz).

Inorganic polyphosphates (polyP) consist of linear chains of orthophosphate units, linked together by high-energy phosphoanhydride bonds. The family of polyP molecules are evolutionarily old biopolymers and are found from bacteria to man. PolyP is exceptional, no other molecule concentrates as much (bio)chemically usable energy as polyP in animals, including humans. Before this discovery, we found that the long-neglected polymer provides ortho-phosphate units, required for bone (hydroxyapatite) synthesis. Hence, polyP is a cornerstone for bone synthesis and repair, especially in higher animals. Besides of its importance for regenerative medicine, especially for the reconstitution of osteo-articular impairments/defects, a further imperative property could be attributed the polyP. This polymer is the only extracellular

generator of metabolic energy in the form of ATP. While the mitochondria synthesize ATP in large amounts intracellularly, it is the polyP, which functions as the storage for extracellular ATP. After enzymatic hydrolysis of polyP through the alkaline phosphatase (ALP) the released free energy is partially stored in ADP (during the transition from AMP), which in the second step is up-phosphorylated to ATP by the adenylate kinase (ADK). In turn, the two enzymes ALP and ADK are the bio-catalytic proteins that conserve the released free energy and store it in ATP, especially in the extracellular space. In a proof-of-concept we could demonstrate that polyP is an essential component for human regeneration processes, especially in those regions which are poorly vascularised, like in bone, cartilage and wounds (including chronic wounds).

[7246] *Preliminary Study on the Identification of Diseases by Electrocardiography Sensors' Data*

Rui João Pinto (University of Trás-os-Montes e Alto Douro), Pedro Miguel Silva (University of Trás-os-Montes e Alto Douro), Rui Pedro Duarte (University of Trás-os-Montes e Alto Douro), Francisco Alexandre Marinho (University of Trás-os-Montes e Alto Douro), António Jorge Gouveia (University of Trás-os-Montes e Alto Douro), Norberto Jorge Gonçalves (University of Trás-os-Montes e Alto Douro), Paulo Jorge Coelho (Polytechnic of Leiria), Eftim Zdravevski (University "Ss.Cyril and Methodius" Faculty Of Computer Science and Engineering), Petre Lameski (University "Ss.Cyril and Methodius"), Nuno Garcia (Universidade da Beira Interior) and Ivan Miguel Pires (Instituto de Telecomunicações, Universidade da Beira Interior, and Universidade de Trás-os-Montes e Alto Douro).

An electrocardiogram (ECG) is a simple test that checks the heart's rhythm and electrical activity and can be used by specialists to detect anomalies that could be linked to diseases. This paper intends to describe the results of several artificial intelligence methods created to automate identifying and classifying potential cardiovascular diseases through electrocardiogram signals. The ECG data utilized was collected from a total of 46 individuals (24 females, ages 26 to 90, and 22 males, ages 19 to 88) using a BITalino (r)evolution device and the OpenSignals (r)evolution software. Each ECG recording has a duration of around 60 seconds, where, during 30 seconds, the individuals were in a standing position and seated down during the remaining 30 seconds. The best performance in identifying cardiovascular diseases with ECG data was achieved with the Naive Bays classifier, reporting an accuracy of 81.36%, a precision of 26.48%, a recall of 28.16%, and an F1-Score of 62.80%.

[7286] *Predicting Papillary Renal Cell Carcinoma Prognosis Using Integrative Analysis of Histopathological Images and Genomic Data*

Shaira Kee (Department of Natural Sciences, University of St. La Salle, Bacolod, Philippines), Michael Aaron Sy (Department of Electronics Engineering, University of St. La Salle, Bacolod, Philippines), Samuel Border (J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States), Nicholas Lucarelli (J. Crayton Pruitt Family Department of Biomedical Engineering, University of Florida, Gainesville, FL, United States), Akshita Gupta (Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States), Pinaki Sarder (Division of Nephrology, Hypertension and Renal Transplantation, Department of Medicine, University of Florida), Marvin Masalunga (Medical Education Unit, School of Medicine, Southwestern University - PHINMA, Cebu, Philippines) and Myles Joshua Tan (Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, United States).

Renal cell carcinoma (RCC) is a common malignant tumor of the adult kidney, with the papillary subtype (pRCC) as the second most frequent. There is a need to improve evaluative criteria for pRCC due to overlapping diagnostic characteristics in RCC subtypes. To create a better prognostic model for pRCC, we proposed an integration of morphologic and genomic features. Matched images and genomic data from The Cancer Genome Atlas were used. Image features were extracted using CellProfiler, and prognostic image features were selected using least absolute shrinkage and selection operator and support vector machine algorithms. Eigengene modules were identified using weighted gene co-expression network analysis. Risk groups based on prognostic features were significantly distinct ($p < 0.05$) according to Kaplan-Meier analysis and log-rank test results. We used two image features and nine eigengene modules to construct a model with the Random Survival Forest method, measuring 11-, 16-, and 20-month areas under the curve (AUC) of a time-dependent receiver operating curve. The integrative model (AUCs: 0.877, 0.769, and 0.811) outperformed models trained with eigengenes alone (AUCs: 0.793, 0.748, and 0.772) and morphological features alone (AUCs: 0.547, 0.497, 0.483). This suggests that an integrative prognostic model based on histopathological images and genomic features could significantly improve survival prediction for pRCC patients and assist in clinical decision-making.

[7317] *Portable MRI System Based on the Gradient-Free Imaging Technique*

Boguslaw Tomanek (University of Alberta), Aaron Purchase (University of Alberta), Christoper Sedlock (University of Alberta) and Jonathan Sharp (University of Alberta).

The presented portable MRI system is based on the TRASE (Transmit Array Spatial Encoding) imaging method that does not require gradients of the magnetic field to collect 3D MR images. Unlike standard MRI systems, the portable MRI is silent; hardware is inexpensive and light by avoiding a large gradient subsystem, and large and heavy magnet. This new technology provides opportunity for worldwide accessibility of MRI systems. However, construction of the TRASE based system brings challenges, such as the requirements for complex radio-frequency (RF) coils and high duty cycle RF amplifiers. The TRASE principles, the solutions to the challenges and opportunities for new imaging techniques will be explained. In particular, construction of RF coils and a new magnet suitable for TRASE will be presented.

[7436] *ITRAQ-based proteomic analysis reveals potential osteogenesis-promoted role of ATM in strontium-incorporated titanium implant*

Yuzi Xu (Zhejiang University), Yangbo Xu (Zhejiang University) and Fuming He (Zhejiang University).

Objective: Surface modification of implants is one of the research highlights worldwide. Among them, titanium sandblasted-etched surface (SLA) implants have been widely used in clinical treatments due to their excellent osseointegration ability. In our previous studies, we found that strontium-incorporated titanium (SLA-Sr) implants prepared by hydrothermal method showed better osteogenesis-promoted ability compared with SLA implants, but the specific molecular mechanism was still unclear. This study aims to build a "bridge" between traditional experiments and data analysis through bioinformatics methods, to reveal the potential biomolecules and signal pathways that SLA-Sr implants promote osteogenesis. Materials and methods: SLA and SLA-Sr implants were inserted in the metaphysis of tibia in rats, and isotopic labelled relative and absolute quantitative (iTRAQ) techniques were used to screen proteins with potential osteogenesis. Biological functions and signal enrichment pathways of differentially expressed proteins were analysed using Gene Ontology database and Reactome database. The STRING database is used to construct a protein interaction network. Finally, DNA damage repair related genes were selected for validation at mRNA and protein expression levels. Results: In total, 19,343 peptides and 4,280 proteins were identified, among which, about 97% of the peptides were between 6 and 28 amino acid residues, about 98% of the proteins were greater than 10kDa, and over 91% of the proteins were less than 10 peptides. On the third day of implantation, 13 down-regulated proteins and 78 up-regulated proteins were identified in the SLA-Sr group compared with the control group. On the seventh day, 59 down-regulated proteins and 79 up-regulated proteins were detected in the SLA-Sr group. The biological function of the differentially expressed proteins on day 3 was mainly related to actin filament gliding, while the biological function of the differentially expressed proteins on day 7 was mainly related to "mismatch repair" and "DNA damage checkpoint in G1 phase of mitosis", etc. Among them, Atm protein up-regulated on the third day showed a trend of further up-regulation on the seventh day. Analysis of the Reactome signalling pathway showed that Atm was associated with "regulation of TP53 activity after phosphorylation" and "TP53 regulation of transcription of kinases and their excitons". RT-PCR and IHC validation results demonstrated that mRNA and protein expression levels of Atm was higher in SLA-Sr group. Conclusion: SLA-Sr titanium implant could initiate DNA damage repair by activating expression levels of ATM. This study was striving to reveal new faces of better osseointegration and shedding light on the biological function and underlying mechanisms of this important procedure.

[7437] *BCAnalyzer: A semi-automated tool for the rapid quantification of cell monolayer from microscopic images in scratch assay*

Aleksandr Sinitca (St. Petersburg Electrotechnical University ''LETI''), Airat Kayumov (Institute for Fundamental Medicine and Biology, Kazan Federal University), Pavel Zelenikhin (Institute for Fundamental Medicine and Biology, Kazan Federal University), Andrey Porfiriev (Institute for Fundamental Medicine and Biology, Kazan Federal University), Dmitrii Kaplun (St. Petersburg Electrotechnical University ''LETI'') and Mikhail Bogachev (St. Petersburg Electrotechnical University ''LETI'').

The scratch assay is a simple and low-cost approach to evaluate the speed and character of cell migration {\it in vitro}. The principle is based on the online imaging of the ``scratch'' in the cells monolayer being filled with new cells from both edges in real time. Thus, the scratch assay represents a model of cell migration during wound healing and is compatible with imaging of live cells during migration under various conditions. For the quantitative assessment of the scratch area

in microscopic images, we suggest a simple semi-automated two-step algorithm based on the local edge density estimation, which does not require any preliminary learning or tuning, although with a couple of parameters directly controllable by the end user to adjust the analysis resolution and sensitivity, respectively. Using several representative examples of cell lines, we show explicitly the effectiveness of the image segmentation and quantification of the cells monolayer and discuss benefits and limitations of the proposed approach. A simple open-source software tool based on the proposed algorithm with on-the-fly visualization allowing for a straightforward feedback by an investigator without any specific expertise in image analysis techniques is freely available online at https://gitlab.com/digiratory/biomedimaging/bcanalyzer

---

[7451] *Annotation-free Identification of Potential Synteny Anchors*

Karl-Kristian Kaether (Bioinformatics Group, Universitaet Leipzig), Steffen Lemke (Center for Organismal Studies, Universität Heidelberg, Germany) and Peter F. Stadler (Bioinformatics Group, Universität Leipzig).

Orthology assignment between genetic elements lies at the heart of comparative genomics. Current methods primarily rely on se- quence and structural similarity. Both low sequence similarity and the presence of multiple copies limit similarity-based methods. Synteny, i.e., conservation of (relative) genomic location, can help resolve many such cases. The mapping of synteny is based on “synteny anchors”, defined as intervals of genomic locations for which unique orthologs in related genomes can be determined unambiguously. Usually, annotated elements such as protein-coding genes are utilized for this purpose. Here we de- scribe an annotation-free approach and devise a k-mer-based heuristic to identify synteny anchors. To demonstrate the practicality of the ap- proach, we compute and analyze a set of synteny anchors for 25 Drosophila species.

[7471] *Bioinformatics approaches to characterize the Monkeypox virus genomes from cases of the mid-2022 outbreak*

Adrián Muñoz-Barrera (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain), Laura Ciuffreda (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain), Julia Alcoba-Florez (Servicio de Microbiología, Hospital Universitario Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain), Luis A. Rubio-Rodríguez (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain), Héctor Rodríguez-Pérez (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain), Helena Gil-Campesino (Servicio de Microbiología, Hospital Universitario Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain), Diego García-Martínez de Artola (Servicio de Microbiología, Hospital Universitario Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain), Josmar Salas-Hernández (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain), Julia Rodríguez-Núñez (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain), Antonio Íñigo-Campos (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain), Víctor García-Olivares (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain), Oscar Díez-Gil (Servicio de Microbiología, Hospital Universitario Ntra. Sra. de Candelaria, Santa Cruz de Tenerife, Spain), Rafaela González-Montelongo (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain), Agustín Valenzuela-Fernández (Lab. de Inmunología Celular y Viral, U. de Farmacología, Facultad de Medicina, Univ. de La Laguna, La Laguna, Spain), José M. Lorenzo-Salazar (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain) and Carlos Flores (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain).

1 Introduction In early May 2022, the World Health Organization declared community transmission of monkeypox (Mpox) due to an outbreak in several countries [1]. Since the start of this Mpox outbreak and as of February 2023, more than 20,000 confirmed cases of Mpox have been reported from 29 EU/EEA countries. In Spain, several cases of Mpox infection have been associated with outbreaks in the Canary Islands [2]. Here we describe different bioinformatics approaches used to characterize the sequence of the first Mpox viral genome isolated from a patient with mild symptoms in the Canary Islands.

2 Materials and methods Viral DNA from five samples (nasopharyngeal swab, lesion crust, and vesicles) isolated from a male adult patient with one week-onset mild symptoms (fever, odynophagia) were subjected to different metagenomics DNA sequencing approaches based on short read (Illumina Inc.) and long read (Oxford Nanopore Technologies) technologies for Mpox sequencing. Two bioinformatics processing approaches were used to obtain the viral genome

sequence: i) by aligning the short reads to a reference (MT903344.1), and ii) using a de novo hybrid assembly from both short-and long reads. The generated sequences in FASTA format were uploaded to Nextclade v.2.11 for clade assignment, mutation calling, and sequence quality checks. Both sequences have been deposited in NCBI GenBank under the accession numbers ON782054 and ON782055.

A full description of the pipelines is available at the GitHub repository: https://github.com/genomicsITER/monkeypox

3 Results Aligning the short reads to a reference provided the best result. Depending on the tools used for alignment and variant identification, between 46 and 67 additional mutations to those observed in the genomic sequences of the outbreak recorded in 2018-2019 were detected [3]. Phylogenetic analysis placed the sequence obtained in the outbreak in the Canary Islands in clade B.1. Furthermore, the closest sequences in the phylogenetic tree were the Mpox sequences published in GenBank detected in Slovenia in the same period, providing further evidence of community transmission of this virus during the mid-2022 outbreak.

4 Conclusions The sequence of the first Mpox viral genome isolated in the Canary Islands, corresponding to clade B.1, which was observed throughout Europe and other non-endemic areas during the 2022 outbreak [4], has been obtained. The results underline the potential of the use of sequencing technologies, and metagenomics in particular, in the diagnosis and epidemiological surveillance of emerging and re-emerging pathogens.

5 Funding Cabildo Insular de Tenerife (CGIEU0000219140 and “Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19”); Instituto de Salud Carlos III (FI18/00230 and CD22/00138) cofunded by European Union (ERDF) “A way of making Europe”; Fundación Canaria Instituto de Investigación Sanitaria de Canarias (PIFIISC 21/37); Agreement with Instituto Tecnológico y de Energías Renovables (ITER) to strengthen scientific and technological education, training, research, development and innovation in Genomics, Personalized Medicine and Biotechnology (OA17/008); European Centre for Disease Prevention and Control (ECDC/HERA/2021/024 ECD.12241); and Convenio Marco de Cooperación Consejería de Educación-Cabildo Insular de Tenerife 2021-2025 (CGIAC0000014697).

References 1. Multi-country monkeypox outbreak in non-endemic countries, https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON385, last accessed 2023/02/20. 2. Illumina whole-genome sequence of Monkeypox virus in a patient travelling from the Canary Islands to France, https://virological.org/t/illumina-whole-genome-sequence-of-monkeypox-virus-in-a-patient-travelling-from-the-canary-islands-to-france/829, last accessed 2023/02/20. 3. Multi-country outbreak of Monkeypox virus: genetic divergence and first signs of microevolution, https://virological.org/t/multi-country-outbreak-of-monkeypox-virus-genetic-divergence-and-first-signs-of-microevolution/806, last accessed 2023/02/20. 4. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus, https://virological.org/t/urgent-need-for-a-non-discriminatory-and-non-stigmatizing-nomenclature-for-monkeypox-virus/853, last accessed 2023/02/20.

[7473] *Web System For The Rehabilitation Of Cognitive Functions In Patients With Traumatic Brain Injury*

Mirta Moran (Biomedical Engineering Research Group - National University of Asunción), César Yegros (Biomedical Engineering Research Group - National University of Asunción) and Eladio Quintana (Biomedical Engineering Research Group - National University of Asunción).

The Traumatic Brain Injury (TBI) is brain damage that occurs suddenly due to a blow to the head or an element that passes through it. Traumatic Brain trauma constitutes one of the main causes of disability in the young population, previous-ly healed, especially as a result of road accidents mainly motorcycle accidents. The physical and cognitive alterations caused by traumatic brain trauma limit the performance in the work, academic and social activities of the affected person. The cognitive sequelae interfere with the daily activities of the individual directly affecting their quality of life and that of their environment; Therefore, neuropsy-chological intervention is necessary through a cognitive rehabilitation program. The Basic Health Indicators 2019 of the Ministry of Public Health and Social Welfare show that the health regions with the highest mortality rate due to Land Transportation Accidents are: President Hayes, Amambay, Concepción and Cor-dillera. Considering the incidence data of land transport accidents, it is necessary to provide post-TBI patients with a care service in the area of neuropsychology, especially for patients residing in the countryside. The main objective of the pro-ject is to develop a system based on Information and Communication Technolo-gies (ICTs) that can be implemented in the platform of the National Telemedicine System of the MSP and BS to facilitate the process of rehabilitation of cognitive functions after Traumatic Brain Injury (TBI)

[7485] *Constructing a Stroke Diagnosis and Prognosis System Based on the BPN Algorithm Using Tc-99m-ECD SPECT images*

Jui-Jen Chen (Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University College of Medicine), Hung-Nien Chang Chien (Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University College of Medicine) and Yen-Hsiang Chang (Chang Gung Memorial Hospital, Kaohsiung Medical Center, Chang Gung University College of Medicine).

The technology of deep learning in artificial intelligence (AI) is increasingly applied to medical image recognition. These applications are focused mainly on the analysis of CT or MRI images and seldom intended to support re-search of nuclear medical images. SPECT is one of the few examination tools that can sensitively reflect abnormalities in the brain at an early stage. In addition to diagnosing cerebrovascular diseases, it can also be used as a tool for prognostic evaluation. Therefore, constructing a stroke image recog-nition system for diagnosis and prognosis is the goal of this study. We col-lected all the Tc-99m ECD SPECT brain images from Kaohsiung Chang Gung Memorial Hospital over a period of five years from 2017 to 2021. A total of 144 medical records that met the ICD-10-Code I60-I69 cerebrovas-cular disease extraction rules were obtained. In the preprocessing of data, noise and defective images were removed. Data augmentation technology was exploited to avoid overfitting and underfitting due to the small amount of data and obtain more new images for higher generalization ability. The back-propagation neural network (BPN) algorithm was adopted to train stroke images and extract important features according to the distribution of blood flows in the brain. The proposed model is compared with the VGG16 through transfer training. It delivers an accuracy of 94.4% (1.3% higher) and a sensitivity of 90.3% (5.4% higher). Its recall rate and F-score reach 90.3% and 94.2% respectively. The ROC curve and average AUC (0.89±0.08) indi-cate that this model has an excellent discrimination capability. Conclusion: Based on the strengths of the BPN algorithm, we construct a stroke image recognition system to support stroke image recognition, assessment of the possibility of a second stroke, and prognostic evaluation. This system can al-so assist physicians in performing a rapid diagnosis and reduce errors. It is hoped that the developed software can be ported to real-world medical sys-tems for testing, so as to connect the theory with practical situations.

[7587] *Deep Learning for Parkinson's Disease Severity Stage Prediction using a New Dataset*

Zainab Maalej (Institut Supérieur de Gestion de Tunis), Fahmi Ben Rejab (Institut Supérieur de Gestion de Tunis) and Kaouther Nouira (Institut Supérieur de Gestion de Tunis).

Parkinson's Disease (PD) is a progressive neurological disorder affecting the Basal Ganglia (BG) region in the mid-brain producing a degeneration of motor abilities. The severity assessment is generally analyzed through Unified Parkinson's Disease Rating Scale (UPDRS) as well as the amount changes noticed in the BG size in Positron Emission Tomography (PET) images. Predicting patients' severity state through the analysis of these symptoms over time remains a challenging task. This paper proposes a Long Short Term Memory (LSTM) model using a new created dataset in order to predict the next severity stage. The dataset includes the UPDRS scores and the BG size for each patient. This is performed by implementing a new algorithm that focuses on PET images and computes BG size. These computed values were then merged with UPDRS scores in a CSV file. The dataset created is fed into the proposed LSTM model for predicting the next severity stage by analyzing the severity scores over time. The model's accuracy is assessed through several experiments and reached an accuracy of 84\% which outperforms the other state-of-the-art method. These results confirm that our proposal holds great promise in providing a visualization of the next severity stage for all patients which aid physicians in monitoring disease progression and planning efficient treatment.

[7653] *Selection Process of Phytochemicals and Efficacy of Thymol, Eugenol and Calcium Ferulate on Heterotrophic Plate Count Bacteria in Water*

Humayun Wali (University of Engineering and Technology, Lahore) and Muhammad Zafar (University of Engineering and Technology, Lahore).

Water samples of Lahore Canal have been tested for log reduction of heterotrophic plate count microorganisms in the presence of three selected phytochemicals: thymol, eugenol and freshly prepared calcium salt of ferulic acid. Thymol results in about 0.9 log10 reduction (at 80 min contact time), 1.2 log10 reduction (at 90 min contact) and 2.6 log10 reduction (contact time: 60 min) for 75, 150 and 300 ppm wt./wt. phytochemical in water sample respectively. Eugenol at 150 and 300 ppm is as good as thymol for log10 reduction, but not at 75 ppm. Calcium ferulate does not significantly reduce the heterotrophic plate count microorganisms at the three tested concentrations. Contact time is found crucial

for optimum reduction of microorganisms. Further studies have been carried out for thymol at 50 ppm. It is observed that thymol results in 0.9 log10 reduction at pH 9.5 and 30oC. At this concentration and at 13oC and 20oC, significant decrease in heterotrophic plate count is not observed at either pH 4.5, 7.0, and 9.5. Contact time is again important for inactivating the microorganisms. Statistical analysis is made to evaluate differences between the phytochemical-time in concentration treatment groups and pH-time in the temperature treatment groups. Two linear models of thymol reduction of heterotrophic plate count microorganisms are developed to study the relationships between the variables responsible for inactivation. Where experimental data is not available, software PASS, GUSAR, EPI Suite and Marvin Sketch are used to predict antimicrobial properties, toxicity and water solubility. Based on these criteria, affordability and aesthetics, the final selection of phytochemicals is made.

[7684] *Osteopontin overexpression synergistically interacts with Aurora kinases overexpression, and is associated with tumor progression, early tumor recurrence, and poor prognosis in hepatocellular carcinoma*

Zhong-Zhe Lin (Natinaol Taiwan University Cancer Center), I-Lun Tsai (Natinaol Taiwan University Cancer Center) and Kuan-Yu Chen (Natinaol Taiwan University Cancer Center).

Background: 1. Hepatocellular carcinoma (HCC) is the leading cause of cancer mortality worldwide. The detailed molecular mechanisms of hepatocarcinogenesis are still not fully understood. 2. The identification of molecular markers related to hepatocarcinogenesis would provide for better management planning and serve as potential therapeutic targets. 3. We previously reported Osteopontin [1], Aurora A [2], and Aurora B [3] are highly expressed in HCC, and the overexpression is closely associated with aggressive tumor phenotypes. Our results showed that Osteopontin, Aurora A, and Aurora B mRNA overexpression, independent of p53/β-catenin mutations, are important molecular markers associated with early tumor recurrence and poor patient prognosis. 4. In this study, we sought to analyze the interactions of these molecular factors, as well as the prognostic significance of the interactions in HCC.

Method: 1. During the period January 1987 through December 1997, 203 surgically resected, primary unifocal HCCs were selected for this study. 2. Age, gender, hepatitis status,α-fetoprotein (AFP) levels, liver cirrhosis, portal venous thrombosis (PVT), tumor size, histology grade, tumor stage, Osteopontin/Aurora A/Aurora B mRNA levels, mutations of the p53 and β-catenin genes, and early tumor recurrence within 12 months after surgery (ETR) were analyzed for their impact on patient survival. 3. Tumor grade was divided into 3 groups: well-differentiated (grade I), moderately differentiated (grade II), and poorly differentiated (grade III-IV). Staging was based on the International Union Against Cancer criteria, with slight modification because HCC tends to spread in the liver via vascular invasion [4]. 4. Osteopontin/Aurora A/Aurora B mRNA levels were measured in 203 HCC and paired non-tumorous liver tissues by reverse transcription-PCR. The mRNA levels were determined by the ratio of signal intensity of target genes to that of S26, and categorized as overexpression if the ratio > 1.0. Mutations of the p53 and β-catenin genes were detected by direct sequencing of exon 2 to exon 11 of p53 and exon 3 of β-catenin. 5. Data analyses were carried out with Statistical Analysis System software (version 9.1; SAS Institute, Inc., Cary, NC). The 2, Fisher's exact test, and log-rank test were used for univariate analysis. Multivariate analyses were conducted for tumor grade and stage by fitting multiple logistic regression models, and time to ETR and time to death were analyzed by fitting multiple Cox's proportional hazards models.

Results: 1. As shown in Table 1 and Figure 1, AFP elevation, PVT, large tumors, advanced tumor grade/stage, overexpression of Aurora A/B, OPN, p53 mutation, and ETR were associated with worse overall survival of patients. Chronic hepatitis C andβ-catenin mutation correlated to better patient survival. 2. OPN overexpression correlated with Aurora A overexpression (Odd ratio [OR], 3.16; P <0.001), Aurora B overexpression (OR, 2.79; P =0.002). OPN overexpression also showed trends to positively correlate to p53 mutation (OR, 1.27; P =0.501) and negatively correlate to β-catenin mutation (OR, 0.84; P =0.804) (Table 2). 3. Multivariate analysis showed that ETR (OR, 5.246; P <0.001), PVT (OR, 1.939; P =0.034) were significant independent risk factors associated with poor patient survival. PVT (OR, 2.59; P =0.01), tumor size larger than 5 cm (OR, 2.527; P =0.01), and p53 mutation (OR, 2.215; P =0.017) were significant independent risk factors by Cox's proportional hazards model for the occurrence of ETR. 4. Overexpression of each of the markers significantly correlated with overexpression of the other two. The coexistence of overexpressions of osteopontin and Aurora A/B were especially associated with worse survival outcomes of the HCC patients (Figure 2).

Conclusions: Overexpressions of osteopontin, Aurora A, and Aurora B are frequently found in HCC, and these three molecular factors had intimate interactions. Overexpressions of osteopontin and Aurora A synergistically led to poor prognosis for HCC patients.

[7881] *Assessing bioinformatic tools for de novo assembly of nanopore sequencing data from human whole-genomes*

Adrián Muñoz-Barrera (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife), Víctor García-Olivares (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife), Luis A. Rubio-Rodríguez (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife), David Jáspez (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife), José M. Lorenzo-Salazar (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife), Rafaela González-Montelongo (Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife) and Carlos Flores (Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife).

Recent advances in third-generation sequencing (TGS) technologies, such as nanopore-based sequencing, have made it possible to obtain long DNA sequences that, in turn, facilitate the de novo assembly of genomes. However, accurate and precise de novo assemblies of large and complex genomes remains challenging and requires robust benchmarking of available tools. In this context, several software tools have been developed for de novo genome assembly using long-read noisy data, though their performance can vary depending on the quality and quantity of input data, as well as other factors such as genome size and complexity. Therefore, the benchmarking of different de novo assembly tools using data from long-read sequencing is essential for improving the accuracy and quality of de novo genome assembly, and to facilitate the discovery of new genomic features and functions in the human genome that may be medically relevant.

[8024] *Quantitative EEG Findings in Outpatients with Psychosomatic Manifestations after COVID-19*

Sergey Lytaev (Saint Petersburg State Pediatric Medical University), Nikita Kipaytkov (Saint Petersburg State Pediatric Medical University) and Tatayna Navoenko (Saint Petersburg State Pediatric Medical University).

EEG is considered an important tool in the diagnostic and treatment process of patients with neurological manifestations of COVID-19, especially with encepha-lopathy, seizures, and status epilepticus. The present research was aimed at quantitative and visual analysis of the EEG of 85 neuropsychiatric outpatients with COVID-19 history and with psychosomatic complaints at the time of the exami-nation. The control group consisted of 35 healthy subjects. Three types of EEG patterns have been established: polymorphic low-frequency activity; low-frequency polymorphic activity with a predominance of delta, theta rhythms; high frequency EEG with a visible dominant of the beta1 range. The correlation index in the alpha range is stable for the EEG in the control group, where in 90% of the subjects the correlation coefficients in the alpha range were more than 0.6. On the contrary, patients have a polymorphic picture, stable indicators with a coefficient of more than 0.6 for all the studied connections, both between the hemispheres and within the hemispheres were registered only in 25% of the subjects. Analysis of the coherence coefficients in patients, on the contrary, shows a higher stability of interhemispheric connections and various options for reducing connections within the hemispheres, which often have a “mirror character”.

[8290] *Color Hippocampus Image Segmentation using Quantum Inspired Firefly Algorithm and Merging of Channel-wise Optimums*

Alokeparna Choudhury (Indian Institute of Information Technology, Kalyani, India), Sourav Samanta (Indian Institute of Information Technology, Kalyani, India), Sanjoy Pratihar (Indian Institute of Information Technology, Kalyani, India) and Oishila Bandyopadhyay (Indian Institute of Information Technology, Kalyani, India).

Color image segmentation is essential for medical image processing to figure out the cells, tissues, lesion areas, etc. The hippocampus is an extension of the temporal lobe of the brain. This area of the brain has been intensively studied for its clinical significance. It is the first and most severely affected structure in neuropsychiatric conditions. Meta-heuristic algorithm-based optimal segmentation is a widely accepted method in the medical domain. In this work, a hybrid method called the quantum-inspired firefly algorithm (QIFA) has been implemented in a multi-core environment to perform the color segmentation of the hippocampus image in a parallel manner by running QIFA on three different channels, Red, Green, and Blue of the input color image and a subsequent merging. The correlation has been considered as the objective function. Finally, a study has been done using various image segmentation evaluation parameters, and the proposed method has been compared to other metaheuristic algorithms. The analysis of the results shows that the method is

effective for medical image segmentation. The speed-up of the technique has also been examined in detail for various image sizes and color levels.

[8322] *Breast Cancer Histologic Grade Identification by Graph Neural Network Embeddings*

Salvatore Calderaro (Department of Mathematics and Computer Science), Giosuè Lo Bosco (University of Palermo - Department of Mathematics and Computer Science), Filippo Vella (Institute for High-Performance Computing and Networking, National Research Council of Italy) and Riccardo Rizzo (Institute for High-Performance Computing and Networking, National Research Council of Italy).

Deep neural networks are nowadays state-of-the-art methodologies for general-purpose image classification. As a consequence, such approaches are also employed in the context of histopathology biopsy image classification. This specific task is usually performed by separating the image into patches, giving them as input to the Deep Model and evaluating the single sub-part outputs. This approach has the main drawback of not considering the global structure of the input image and can lead to avoiding the discovery of relevant patterns among non-overlapping patches. Differently from this commonly adopted assumption, in this paper, we propose to face the problem by representing the input into a proper embedding resulting from a graph representation built from the tissue regions of the image. This graph representation is capable of maintaining the image structure and considering the relations among its relevant parts. The effectiveness of this representation is shown in the case of automatic tumor grading identification of breast cancer, using public available datasets.

[8336] *Structural Analysis of RNA-Binding Protein EWSR1 Involved in Ewing's Sarcoma through Domain Assembly and Conformational Molecular Dynamics Studies*

Saba Shahzadi (Nationwide Children's Hospital), Mubashir Hassan (Nationwide Children's Hospital) and Andrzej Kloczkowski (Nationwide Children's Hospital).

Proteins are key regulatory modulators in the prognosis and therapeutics for multiple sarcomas through activating and deactivating the downstream signaling cascades. Current work focuses on the structural and conformational interactive behavior of EWS RNA-Binding Protein 1 (EWSR1) with RNA involved in Ewing's sarcoma. The comparative modeling approach was employed to predict the structural models of EWSR1 domains separately, and the Domain Enhanced MOdeling (DEMO) server for multidomain protein structure assembly was used to predict the structure composed of all three domains. Furthermore, RNA motifs binding to EWSR1 were predicted and the 3D model of RNA bound to EWSR1 was built by using MC-Fold. Protein-RNA dockings to check the conformational interactions between EWSR1 and RNA were studied with HNADOCK. Moreover, Molecular Dynamics simulations were performed to check the stability of EWSR1-RNA complex by computing root mean square deviation and fluctuation (RMSD/F), radius of gyration (Rg) and solvent accessible surface area (SASA). The generated results explored the structural insights on EWSR1domains and their interactive binding with RNA. Based on computational assessments, it has been concluded that certain structural and dynamical features of EWSR1-RNA complex could be used as a target in future development of drugs against Ewing's sarcoma.

[8452] *Principaux facteurs de prise de poids pendant le confinement COVID-19 dans une population Marocaine : une étude transversale*

Nourelyakine Lakhdar (FACULTE DES SCIENCES - IBN TOFAIL), Driss Lamri (FACULTE DES SCIENCES - IBN TOFAIL) and Moulay Ouahidi (FACULTE DES SCIENCES - IBN TOFAIL).

Introduction : L'obésité est très préoccupante en tant que comorbidité de la maladie à coronavirus 2019 (COVID-19). Nous avons examiné les facteurs associés à la prise de poids chez les élèves marocain pendant la pandémie de COVID-19. Méthodologie : Nous avons mené une enquête sur terrain auprès de 500 élèves (315 garçons et 185 filles âgés de 12 à 20 ans) en novembre 2022. Une analyse de régression logistique multivariable a été effectuée pour évaluer les facteurs associés à la prise de poids. L'analyse a été ajustée en fonction du sexe, de l'âge, de l'humeur dépressive, de l'anxiété, des repas au restaurant, des repas tardifs, de l'exercice, des troubles du sommeil, des habitudes de repas, de l'image corporelle subjective, des comorbidités, du fait de vivre seul et du revenu de famille. Résultats : Après ajustement pour tenir compte des variables confusionnelles, les probabilités de gain de poids ont augmenté dans le groupe âgé de 12 à 16 ans par rapport au groupe âgé de 17 à 20 ans (1,82 ; intervalle de confiance [IC] à 95 %, 1,01 à 3,32). Les femmes étaient plus associées au risque de gain de poids que les hommes. Les chances de gain de poids ont augmenté dans le groupe

d'absence d'exercice par rapport au groupe d'exercice (4,89 ; IC à 95%, 3,09-7,88). Les probabilités de gain de poids ont augmenté dans les groupes de repas à l'extérieur et de fin de soirée par rapport à ceux des groupes ne mangeant pas à l'extérieur et n'ayant pas de repas tard le soir. Les personnes qui regardaient un écran pendant 3 à 6 heures par jour étaient plus associées au risque de prise de poids que celles qui regardaient rarement un écran. Les chances de gain de poids ont augmenté chez les participants qui se considéraient obèses par rapport à ceux qui ne se considéraient pas obèses. Conclusion : Une alimentation saine et une activité physique régulière ont tendance à être la meilleure approche pour réduce l'obésité, un facteur de risque de COVID-19.

[8458] *Optimizing Variant Calling for Human Genome Analysis: A Comprehensive Pipeline Approach*

Miguel Pinheiro (iBiMED, Department of Medical Sciences, University of Aveiro, Portugal), Jorge Miguel Silva (IEETA, DETI, LASI, University of Aveiro, Portugal) and José Luis Oliveira (IEETA, DETI, LASI, University of Aveiro, Portugal).

The identification of genetic variations in large cohorts is a critical issue to identify patient cohorts, disease risks, and to develop more effective treatments. To help this analysis, we improved a variant calling pipeline for the human genome using state-of-the-art tools, including GATK (Hard Filter/VQSR) and DeepVariant. The pipeline was tested in a computing cluster where it was possible to compare Illumina Platinum genomes using different approaches. Moreover, by using a secure data space we provide a solution to privacy and security concerns in genomics research. Overall, this variant calling pipeline has the potential to advance the field of genomics research significantly, improve healthcare outcomes, and simplify the analysis process. Therefore, it is critical to rigorously evaluate these pipelines' performance before implementing them in clinical settings.

[8476] *Modelling of Anti-Amyloid-Beta Therapy for Alzheimer's Disease*

Swadesh Pal (MS2Discovery Interdisciplinary Research Institute, Wilfrid Laurier University, Waterloo) and Roderick Melnik (MS2Discovery Interdisciplinary Research Institute, Wilfrid Laurier University, Waterloo).

A healthy brain clears different types of debris with the help of specialized glial cells. These cells contiguously tile the entire central nervous system (CNS), exert many essential complex functions in the healthy CNS, and maintain a healthy balance in the brain. However, over age, these cells fail to control the healthy balance of the proteins and cause different neurodegenerative diseases, one of which is Alzheimer's disease (AD). In AD, insoluble amyloid-beta plaques accumulate in the extracellular space along with neurofibrillary tangles (NFTs) inside the brain cells. In this paper, we have developed a model and have studied the accumulation of amyloid-beta plaques and NFTs along with an anti-amyloid-beta therapy applied in the treatment of the disease. Based on these studies, we have demonstrated the dynamics of the modelling therapy such that the drug helps clear a subsequent amount of amyloid-beta plaques in each dose. Numerical simulations have been used to show different long-term outcomes of the model. To further analyze the disease progression in the brain and its treatment, we have integrated brain connectome data in the network model as part of our developed modelling framework.

[8515] *Modelling the survival kinetics of Salmonella spp. on the surface of ripened raw milk cheese during storage at different temperatures*

Adriana Łobacz (University of Warmia and Mazury in Olsztyn) and Justyna Zulewska (University of Warmia and Mazury in Olsztyn).

The aim of this study was to determine the survival kinetics of Salmonella enterica subsp. enterica in ripened raw milk cheese. Cheese samples inoculated with S. enterica subsp. enterica were stored at 5, 15and 25 °C and analysed in terms of physico-chemical and microbiological characteristics. Three primary models were used to estimate the kinetic parameters of S. enterica subsp. enterica. The secondary Arrhenius model was used to establish the relationship between temperature and parameter a of the Weibull model. Additionally, prediction of S. enterica subsp. enterica survival as a function of storage temperature was made. S. enterica subsp. enterica growth was inhibited during storage, and bacteria survived for an extensive period of time at high number (60 day at 5 °C, 26 day at 25 °C). The storage temperature significantly influenced the inactivation rate of Salmonella in raw milk ripened cheese and proceeded faster at 25 °C compared to remaining storage temperature. Obtained results suggest that contamination by Salmonella in raw milk cheese might result in residual risk.

[8536] *Antimicrobial ceramic materials for biomedical implants*

Julietta V. Rau (Istituto di Struttura della Materia, Consiglio Nazionale delle Ricerche (ISM-CNR), Rome, Italy).

A significant increase of biomedical implant demand is a dominating trend due to the ageing of population. Nowadays, the main requests for implants are their good osteointegration and long term stability, possibly including the host bone regeneration process. For these purposes, metallic implants are coated with various biomimetic functional ceramic materials, substantially improving the properties of metal implant substrate by providing a proper bone-material interface and leading to a better implant integration into the surrounding bone tissue. In the present investigation, recent results for materials characterized by multiple properties, designed as scaffolds and coatings on titanium and on biodegradable metal alloy implants, were obtained. In the case of biodegradable implants, the focus point is the control of their corrosion resistance and bioactivity characteristics. The developed innovative materials were mainly composed of multi-substituted calcium phosphates (hydroxyapatite and tricalcium phosphate) and bioactive glasses, both groups containing trace ions with therapeutic functions, triggering the natural bone tissue response and possessing antibacterial properties. This letter is a challenging issue, which can help to avoid the massive use of antibiotics. In this work, the design and synthesis of antimicrobial materials and the development of antimicrobial surfaces was carried out and their comprehensive characterization was performed. Such properties as structural, morphological, and mechanical features, wetting contact angle, behaviour in model media, etc. were tested. The results for ion multidoped calcium phosphate bioceramics were obtained. Moreover, in vitro bioactivity, cell and microbiology tests data focused on material-cell interactions were derived. The obtained results allowed us to conclude that nanostructured antimicrobial materials developed in this work are promising for new strategies in tissue replacement and regeneration, ensuring the required structural, chemical, morphological and mechanical characteristics and providing a controlled release of active principles, improving the long term stability of implants and providing a microbe free implant surface and improved performances of biomedical implant devices.

[8602] *A pilot study of neuroaesthetics based on the analysis of electroencephalographic connectivity networks in the visualization of different dance choreography styles*

Almudena González (Departamento de Historia del Arte y Filosofía, Universidad de La Laguna), Jose Melendez (Departamento de Educación Física y Salud del ISEF, CURE sede Maldonado de la Universidad de la República) and Julian J González (Departamento de ciencias médicas básicas, Unoversidad de La Laguna, Tenerife).

Neuroaesthetics allows us to understand how the brain works in different artistic languages and, therefore, to broaden the knowledge of our aesthetic judgments. The present pilot study is an interdisciplinary work that aims to differentiate different aesthetic dance choreography styles and to demonstrate the influence of training, learning, enculturation and familiarization of these styles on their brain perception by means of neurophysiological measurements of EEG signals and neural connectivity network analysis techniques. To this end, EEGs of non-expert dancers are recorded while viewing two fragments (film clips) of classical and modern dance and during other control conditions. Measures of functional connectivity between recorded regions are obtained from phase synchronization measurements between pairs of EEG signals in each EEG frequency band (FB). The responses of each FB are evaluated from indices obtained from models of EEG connectivity networks -graphs and connectomes- constructed from graph theory and network-based statistics (NBS) in a global and local context. Thus, significant alterations -in some of the indices- are observed between different contrasts and conditions in certain areas and specific EEG connections -mainly in the Alpha band-. These first results, therefore, suggest the usefulness of this neurosthetic experimental paradigm. On the other hand, these neuroaesthetic procedures may be of special interest in biomedicine because they provide knowledge about different languages that can be applied in therapies and treatments.

[8632] *Analyzing dose parameters of Radiation Therapy Treatment Planning and Estimation of Second Cancer Risks*

Irine Khomeriki (Georgian Technical University), Lily Petriashvili (Georgian Technical University), Maia Topeshashvili (Institute of Clinical Medicine) and Tamar Lominadze (Georgian Technical university).

The number of cancer patients receiving radiotherapy is increasing every year, but fortunately, cancer is no longer considered a death sentence. Advanced technologies such as Intensity Modulated Radiotherapy (IMRT) and Volumetrically Modulated Arc Therapy (VMAT) use modulated intensity beams to achieve lower doses to critical organs, while maintaining target dose coverage. This results in better clinical outcomes, with the same or even lower acute

toxicities and significantly increased long-term survival rates. Radiation therapy is a common treatment for cancer that uses high-energy radiation to destroy cancer cells. The goal of radiation therapy is to target cancer cells while minimizing damage to surrounding healthy tissues. Traditional radiation therapy techniques, such as 3D conformal radiotherapy (3DCRT), deliver radiation in fixed beams that pass through normal tissues to reach the tumor, leading to unwanted side effects. However, with the advancements in technology, radiation therapy has become more precise and effective. Intensity Modulated Radiotherapy (IMRT) and Volumetrically Modulated Arc Therapy (VMAT) are two such advanced techniques that use computer-controlled beams of varying intensity to target the tumor from different angles, while sparing the surrounding healthy tissues. These techniques allow for more accurate and precise delivery of radiation, resulting in lower doses to critical organs and significantly reducing the risk of side effects. The use of these advanced techniques has led to better clinical outcomes and significantly increased long-term survival rates. Patients undergoing radiation therapy with IMRT and VMAT have experienced the same or even lower acute toxicities, such as fatigue, skin irritation, and nausea, than those treated with traditional radiation therapy techniques. This has allowed patients to maintain a better quality of life during and after treatment. However, while modulated techniques spare surrounding tissue, they can increase the low-dose volume, leading to an increased risk of second malignancies. As a result, estimating the risk of secondary cancer has become a major concern and a critical aspect of comparative treatment planning for long-term survival rates. In this study, we conducted a retrospective analysis of dosimetry parameters for different treatment planning techniques (3D Conformal Radiotherapy (3DCRT), IMRT, and VMAT), evaluated the risk of these modalities for developing a solid second cancer using the concept of organ equivalent dose (OED) within the framework of Biological Effects of Ionizing Radiation (BEIR) VII cancer incidence models, and made a decision about the most appropriate treatment technique for a specific tumor site based on disease control rate, secondary cancer risk, and toxicity.

[8802] *Using Digital Biomarkers for Objective Assessment of Perfusionists' Workload and Acute Stress during Cardiac Surgery*

Roger Daglius Dias (Harvard Medical School), Lauren Kennedy-Metz (Roanoke College), Rithy Srey (Veterans Affairs Boston Healthcare System,), Geoffrey Rance (Cape Cod Healthcare), Mahdi Ebnali (Harvard Medical School), David Arney (Harvard Medical School), Matthew Gombolay (Georgia Institute of Technology) and Marco Zenati (Harvard University).

The cardiac operating room (OR) is a high-risk, high-stakes environment inserted into a complex socio-technical healthcare system. During cardiopulmonary bypass (CPB), the most critical phase of cardiac surgery, the perfusionist has a crucial role within the interprofessional OR team, being responsible for optimizing patient perfusion while coordinating other tasks with the surgeon, anesthesiologist, and nurses. The aim of this study was to investigate objective digital biomarkers of perfusionists' workload and stress derived from heart rate variability (HRV) metrics captured via a wearable physiological sensor in a real cardiac OR. We explored the relationships between several HRV parameters and validated self-report measures of surgical task workload (SURG-TLX) and acute stress (STAI-SF), as well as surgical processes and outcome measures. We found that the frequency-domain HRV parameter HF relative power - FFT (%) presented the strongest association with task workload (correlation coefficient:-0.491, p-value: 0.003). We also found that the time-domain HRV parameter RMSSD (ms) presented the strongest correlation with perfusionists' acute stress (correlation coefficient:-0.489, p-value: 0.005). A few workload and stress biomarkers were also associated with bypass time and patient length of stay in the hospital. The findings from this study will inform future research regarding which HRV-based biomarkers are best suited for the development of cognitive support systems capable of monitoring surgical workload and stress in real time.

[8985] *GPU Cloud architectures for bioinformatic applications*

Antonio Maciá-Lillo (Department of Computer Science Technology and Computation, University of Alicante), Tamai Rodríguez (Department of Computer Science Technology and Computation, University of Alicante), Higinio Mora (Department of Computer Science Technology and Computation, University of Alicante), Antonio Jimeno-Morenilla (Department of Computer Science Technology and Computation, University of Alicante) and Jose-Luís Sánchez-Romero (Department of Computer Science Technology and Computation, University of Alicante).

The world of computing is constantly evolving. The trends that are shaping today's applications are Cloud computing and GPU computing. These technologies allow bringing high performance computations to low power devices, when using a computing outsourcing architecture. Following the trend, bioinformatic applications are looking to take advantage of these paradigms, but there are challenges that have to be solved. Data that these applications work one is usually

sensible and has to be protected. Also, GPU usage in Cloud architectures currently presents inefficiencies. This paper makes a review of the characteristics of Cloud computing outsourcing architectures, including the security aspects, and GPU usage for these applications. The proposed architecture includes GPU devices and tries to make efficient use of them. The experiments show that it has the opportunity to increase parallelism and reduce context switching costs when running different applications concurrently on the GPU.

[9048] *Eukaryotic topoisomerases of type IIA: cytoplasmic proteins?*

Jiří Červeň (Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava, Czech Republic), Martin Bartas (Department of Biology and Ecology, Faculty of Science, University of Ostrava) and Petr Pečinka (Department of Biology and Ecology, Faculty of Science, University of Ostrava).

Eukaryotic topoisomerases are key nuclear DNA-binding proteins involved in DNA replication, transcription, recombination, and chromosome segregation. In humans, we can distinguish between six topoisomerases: two of type IB (TOP1 and mitochondrial TOP1MT), two of type IIA (TOP2A and TOP2B), and two of type IA (TOP3A and TOP3B). In 2003, the nuclear export signal (NES) sequence present in human TOP2A and TOP2B was described, indicating that these proteins can be actively translocated from the nucleus to the cytoplasm. However, the cytoplasmic roles of TOP2A and TOP2B are still largely unclear and controversial. In our analysis, we inspected thousands of TOP2A and TOP2B protein sequences across the Eukaryotic domain of life and found that their NES sequences are highly evolutionarily conserved, indicating their functional importance. In addition, based on our current molecular docking results, human TOP2A and TOP2B have been predicted to have a significantly higher binding potential to so-called Z-RNA (which is a double-stranded left-handed form of RNA), in comparison to classical right-handed B-DNA. From this point of view, human TOP2A and TOP2B could theoretically bind cytoplasmic Z-RNA, either human or viral origin. Our data about possible cytoplasmic TOP2B localization are further supported by freely-accessible datasets from the Human Protein Atlas, particularly immunofluorescent staining of human cancer cell lines A-431, U-251MG, and U2OS.

[9284] *Investigating the Dynamics of Cancer Evolution: Variant Allele Frequency Patterns and Model Limitations*

Paweł Kuś (Silesian University of Technology) and Marek Kimmel (Rice University).

The widespread use of Next Generation Sequencing methods has made cancer genomic data more accessible, leading to insights into the molecular mechanisms driving individual cancers. Despite these advancements, our understanding of cancer evolution and the roles of mutagenesis and natural selection remains incomplete. Many existing methods analyze Variant Allele Frequency (VAF) spectra and rely on the model of exponential tumor growth. Under this model, neutral mutations form power-law-shaped neutral tails with a power coefficient of 2, while selected mutations manifest as binomial components. In the real data, this neutral tail is often malformed due to the removal of low-frequency variants. In this study, we developed a model-fitting approach that accounts for the incompleteness of the data and used it to analyze the 76 paired tumor samples. Our results revealed differences between the primary and secondary tumors, including variations in reduced mutation rates and the optimal coefficients of the power-law component of the model. In many cases, the optimal power coefficient deviated from 2, suggesting the presence of competing micro-clones, or invalidity of popular model assumptions such as the exponential population growth model or constant mutation rate. Our findings highlight the need for further research into the mechanisms underlying cancer evolution, including the analysis of multi-sample datasets to better understand the complex interplay of mutagenesis and natural selection in tumor development.

[9307] *Improved Long-term Forecasting of Emergency Department Arrivals with LSTM-based Networks*

Carolina Miranda-Garcia (UNIVERSIDAD DE ALCALA (UAH)), Alberto Garces-Jimenez (UNIVERSIDAD DE ALCALA (UAH)), Jose Manuel Gomez-Pulido (UNIVERSIDAD DE ALCALA (UAH)) and Helena Hernandez-Martinez (UNIVERSIDAD DE ALCALA (UAH)).

Patient admission to Emergency Departments suffers from a great variability. This makes the resource allocation difficult to adjust, resulting in an inefficient service. Several studies have addressed this issue with machine learning's regressors, time series analysis. This research proposes the use of improved recurrent neural networks that consider the dynamic nature of the data, introducing contextual variables that allow improving the predictability. Another important

requirement from ED's administration is to have a wider predicting horizon for short- and long-term resource allocations. The results obtained using the data from one single Hospital in Madrid confirm that the use of deep learning with contextual variables improve the predictability to 6% MAPE for seven days and four months forecasts. As future research lines, the influence of special events, such as seasonal epidemics, pollution episodes, sports or leisure events, as well as the extension of this study to different types of hospitals' emergency departments

---

[9396] *Target NGS data analysis identifies the haplotype in LRRK2 gene as a potential risk factor for endemic parkinsonism*

Radek Vodicka (University Hospital and Palacky University), Kristyna Kolarikova (University Hospital and Palacky University, Olomouc), Petr Kanovsky (University Hospital and Palacky University, Olomouc) and Radek Vrtel (University Hospital and Palacky University, Olomouc).

Parkinson's disease and parkinsonism are relatively common neurodegenerative disorders. This study aimed to assess potential genetic risk factors of haplotypes in genes associated with parkinsonism in a population in which endemic parkinsonism and atypical parkinsonism haverecently been found. The genes ADH1C, EIF4G1, FBXO7, GBA, GIGYF2, HTRA2, LRRK2, MAPT, PARK2, PARK7, PINK1 PLA2G6, SNCA, UCHL1, and VPS35 were analyzed in 62 patients (P) and 69 age-matched controls from the researched area (C1). Variants were acquired by high-throughput sequencing using Ion Torrent workflow. As another set of controls, the whole genome sequencing data from 100 healthy non-related individuals from the Czech population were used (C2); the results were also compared with the Genome Project data (C3). We observed shared findings of four intron (rs11564187, rs36220738, rs200829235, and rs3789329) and one exon variant (rs33995883) in the LRRK2 gene in six patients. A comparison of the C1–C3 groups revealed significant differences in haplotype frequencies between ratio of 2.09 for C1, 1.65 for C2, and 6.3 for C3, and odds ratios of 13.15 for C1, 2.58 for C2, and 7.6 for C3 were estimated. The co-occurrence of five variants in the LRRK2 gene (very probably in haplotype) could be an important potential risk factor for the development of parkinsonism, even outside the recently described pedigrees in the researched area where endemic parkinsonism is present.

[9490] *A Platform for the Radiomic Analysis of Brain FDG PET Images: Detecting Alzheimer's disease*

Ramin Rasi (Biomedical Engineering Institute of Bogazici University) and Albert Guvenis (Biomedical Engineering Institute of Bogazici University).

The objective of this work is to present a radiomics-based platform (RAB-PET) and method to detect Alzheimer's disease (AD) non-invasively using 18FDG-PET images. Radiomic analysis allows the identification of regional features that serve to predict the presence or characteristics of diseases using images as data. We first, used the FastSurfer a deep learning-based toolbox to segment the whole brain into 95 classes by the utilization of the DKT-atlas. Then the PyRadiomics toolbox was used to extract features from 18FDG-PET scans. After preprocessing, the features were subject to a selection process by making use of eight different methods, namely, ANOVA, PCA, Chi-square, LASSO, Recursive Feature Elimination (RFE), Feature Importance (FI), Mutual Information (MI), and Recursive Feature Addition (RFA). Finally, in order to classify the selected features by feature selection methods, we implemented nine different classifier methods, namely, Gradi-entBoosting (GB), RandomForest (RF), DecisionTree (DT), GaussianNB (GNB), GaussianProcess (GP), MLP, QuadraticDiscriminantAnalysis (QDA), AdaBoost (AB), and KNeighbors (KNN) on selected feature subsets. All data (scans and clinical examination results) were obtained from the AD Neuroimaging Initiative (ADNI) database. The RF classifier with 100 iterations on features obtained with the LASSO algorithm yielded an area under the curve of AUC=0.976 with a 95% confidence interval of 0.93-0.98 based on 30% independent test data. We conclude that a platform for radiomic analysis can serve as a potential method for deducing accurate information on brain diseases such as Alzheimer's disease non-invasively using 18FDG-PET images. Further studies are underway to extend this work by studying the association between the set of features and several characteristics of the Alzheimer's disease.

[9515] *Assessing Temporal Stability of Heart Rate Variability Features for Predicting Adverse Cardiovascular Events in Hypertensive Patients*

José María López Belinchón (SIDIS Research Group, Department of Mathematics, UCLM.), Miguel Á. López (SIDIS Research Group, Department of Mathematics, IMACI, UCLM.) and Raúl Alcaraz (Research Group in Electronic, Biomedical and Telecommunication Engineering, UCLM, 16071 Cuenca, Spain.).

Cardiovascular and cerebrovascular diseases are the leading causes of premature death and chronic disability worldwide. Prevention is crucial in managing these chronic and complex diseases, and early identification of high-risk population groups is essential. Recent studies suggest that early diagnosis and intervention can be helpful in preventing premature mortality associated with these diseases. Hypertension, a serious medical condition that affects 1.3 billion people, increases the risk of cardiovascular diseases. Early recognition of symptoms in high-risk subjects is crucial to reduce associated morbidity and mortality. Machine learning classifiers have been used to predict major and fatal cardiovascular events from hypertensive patients by combining common heart rate variability (HRV) features. However, the reproducibility of HRV-based indices from different daytime periods has not been addressed in this context, which is the main goal of this study.

This study used the SHAREE database to analyze the reproducibility of HRV-based features. Twelve consecutive 5 minute-length electrocardiogram (ECG) intervals were selected from each of the 139 hypertensive patients, 17 of whom suffered from fatal cardiovascular events within a year of follow-up. A McNemar test was applied to compare classification differences obtained by the HRV features computed from the different ECG segments. The study found that different classifications were mainly observed in features such as QSE, SPE, MSPE, and LZ2 for consecutive ECG intervals within the same patient. These outcomes suggest a very different potential of HRV-based features to anticipate fatal events in hypertensive patients.

The selection of different ECG intervals and HRV metrics could impact the prediction of fatal cardiovascular events, and both aspects should be thoroughly analyzed in future works. This study highlights the importance of reproducibility and consistency in the use of HRV-based indices for early identification of high-risk patients with hypertension, which is a crucial step towards prevention and improved management of cardiovascular and cerebrovascular diseases.

[9523] *Cyclical learning rates (CLR's) for improving training accuracies and lowering computational cost*

Shrikant Pawar (Claflin University), Aditya Stanam (University of Iowa), Anand Narayanan (Yale) and Rushikesh Chopade (Indian Institute of Technology).

Prediction of different lung pathologies using chest X-ray images is a challenging task requiring robust training and testing accuracies. In this article, one-class classifier (OCC) and binary classification algorithms have been tested to classify 14 different diseases (atelectasis, cardiomegaly, consolidation, effusion, edema, emphysema, fibrosis, hernia, infiltration, mass, nodule, pneumonia, pneumothorax and pleural-thickening). We have utilized 3 different neural network architectures (MobileNetV1, Alexnet, and DenseNet-121) with four different optimizers (SGD, Adam, and RMSProp) for comparing best possible accuracies. Cyclical learning rate (CLR), a tuning hyperparameters technique was found to have a faster convergence of the cost towards the minima of cost function. Here, we present a unique approach of utilizing previously trained binary classification models with a learning rate decay technique for re-training models using CLR's. Doing so, we found significant improvement in training accuracies for each of the selected conditions. Thus, utilizing CLR's in callback functions seems a promising strategy for image classification problems

---

[9612] *Towards ASSURED Diagnostics Using Paper-Microfluidic Integrated Chemiresistor Biosensor Arrays*

Ashok Mulchandani (University of California, Riverside).

Availability of affordable, rapid, sensitive, specific biosensors are critical to prompt disease diagnosis, especially in developing areas with low resources. However, traditional technologies to detect disease biomarkers such as colorimetric lateral flow assays and polymerase chain reaction tests, suffer from low sensing performance, high operation complexity, high cost, and long assay time. Therefore, it is imperative to develop an affordable biosensing platform with high sensing performances and high user-friendliness to meet the WHO ASSURED criteria (affordable, sensitive, specific, user-friendly, rapid and robust, equipment-free, and deliverable to end-users) for diagnostic tools in resource-limited areas. In this presentation I will present our recent work on the development of the paper-based microfluidic chemiresistive biosensors employing single-walled carbon nanotubes (SWCNTs) and specific bioreceptors for the detection of protein and nucleic acid biomarkers in various biological samples for diagnostics in resource-limited settings.

[9647] *Detecting Intra Ventricular Haemorrhage in Preterm Neonates using LSTM Autoencoders*

Idris Muniru (Stellenbosch University), Jacomine Grobler (Stellenbosch University) and Lizelle Van Wyk (Stellenbosch University).

The neonatal period is characterized by numerous physiological adaptations for extra-uterine life, during which newborns are vulnerable to various diseases and disorders. Preterm neonates (PN), those born before 37 weeks gestation, are particularly vulnerable, especially in the first week of life. One of the most common complications in premature neonates is intraventricular hemorrhage (IVH), which occurs in approximately 15-20% of cases and can lead to severe neurological complications such as cerebral palsy, developmental delays, and cognitive impairments. Early detection and intervention are critical to prevent these long-term consequences. Non-invasive cardiac output monitors (NICOM) have been widely accepted for monitoring hemodynamic parameters in the neonatal intensive care unit (NICU) and have generated a wealth of data. However, further exploration of their predictive tendencies with regard to IVH is needed. This study aims to evaluate the potential of deep learning models to improve the early identification and prevention of IVH in preterm neonates using NICOM parameters. The results suggest that while the models performed reasonably well, there is room for improvement, particularly in terms of reducing false alarms. Nonetheless, this study represents a significant step towards developing a non-invasive, accurate, and timely method for monitoring and preventing IVH in preterm neonates, particularly in low-resource settings.

[9662] *Enabling real-time analysis of nanopore 16S rRNA sequencing data with NanoRTax*

Héctor Rodríguez Pérez (Research Unit, Hospital Universitario La Candelaria), Laura Ciuffreda (Research Unit, Hospital Universitario La Candelaria) and Carlos Flores (Research Unit, Hospital Universitario La Candelaria).

Rapid characterization of bacterial communities is important for the diagnosis of bacterial infections and treatment monitoring based on microbiome changes over time. Nanopore sequencing of full-length 16S rRNA gene provides improved taxonomic classification compared to that offered by short-read technologies. However, the lack of adequate bioinformatic pipelines has limited the usage of this technology in clinical settings. We present NanoRTax, a real-time analysis pipeline of nanopore 16S rRNA amplicon sequencing data that integrates all algo-rithms and dependencies in a conda environment or Docker container. The pipe-line includes state-of-the-art classifiers for real-time analysis and a web applica-tion for the visualization of results on the fly and as aggregated data at the end of the run.

[9753] *Measurement Of Acute Pain In The Pediatric Emergency Department Through Automatic Detection Of Behavioral Parameters: A Pilot Study*

Letizia Bergamasco (Fondazione LINKS), Marco Gavelli (Fondazione LINKS), Carla Fadda (Scuola di Specializzazione in Pediatria, Università degli Studi di Torino), Emilia Parodi (S.C. Pediatria e Neonatologia, A.O. Ordine Mauriziano di Torino), Claudia Bondone (S.C. Pediatria d'Urgenza, Ospedale Infantile Regina Margherita – A.O.U. Città della Salute e della Scienza di Torino) and Emanuele Castagno (S.C. Pediatria d'Urgenza, Ospedale Infantile Regina Margherita – A.O.U. Città della Salute e della Scienza di Torino).

Acute pain is a frequent symptom in children who access the Emergency Department (ED). Its measurement through validated tools compatible with the time of triage is essential to develop the most appropriate pain-relieving strategy. The algometric scales that can be used in children in whom self-assessment is not possible are based on the evaluation of behavioral and physiological parameters. However, the actual use of algometric scales in the ED is scarce due to environmental factors, heterogeneity of the scales and lack of training, thus making automated pain assessment desirable. In this study, we propose a camera-based system to provide an objective and contactless pain assessment in children aged less than 3 years, through the automatic detection of behavioral parameters from video recordings. To investigate the feasibility of its usage in the ED environment, we collected video recordings of healthy children aged 3-36 months admitted to the ED with acute pain as the main or accompanying symptom, while pain was measured by a healthcare professional according to the Face, Legs, Activity, Cry, and Consolability (FLACC) pain scale. For the recorded videos, we compared the scores for the items Face (F), Legs (L) and Activity (A) given by the operator with the ones given by our system, analyzing the potentiality and limitations of our approach. By showing that automatic pain assessment in young children in the ED could integrate human evaluation to make it easier and faster, without substituting it, we provide the basis for further research in this field.

[9807] *Gait Asymmetry Evaluation using FMCW Radar in Daily Life Environments*

Shahzad Ahmed (Hanyang University), Yudam Seo (Hanyang University) and Sung Ho Cho (Hanyang University).

Gait analysis plays a crucial role in medical diagnostics due to its ability to determine and quantify the patient's physical abilities and limitations. Unlike the other competing sensors, radar is capable of measuring human gait in non-contact fashion. In this paper, we present the extraction of six different gait parameters using Frequency Modulated Continuous Wave (FMCW) radar within five steps. The range-time and Doppler-time information is used to extract the parameters. The rangetime information of FMCW radar yields the walking duration, walking time, and average walking velocity whereas, the velocity-time information yields pause time during walking, inter-step distance variation, and inter-step time variations. An Inertial Measurement Unit (IMU) is deployed as a ground-truth reference sensor to track the gait movement and a high correlation is found between radar and the reference sensor. Finally, as a use case example, gait parameters analysis is performed to detect asymmetric gait movement. Symmetric and asymmetric walking data is collected with radar and features analysis is performed which suggests that inter-step time and velocity variations contributes greatly in asymmetry detection.

[9863] *The use of H-Scan ultrasound imaging to assess the re-sponse of breast cancer patients to neoadjuvant chemo-therapy*

Hanna Piotrzkowska Wróblewska (Institute of Fundamental Technological Research Polish Academy of Sciences), Katarzyna Dobruch Sobczak (Maria Skłodowska-Curie National Research Institute of Oncology), Ziemowit Klimonda (Institute of Fundamental Technological Research Polish Academy of Sciences), Piotr Karwat (Institute of Fundamental Technological Research Polish Academy of Sciences) and Jerzy Litniewski (Institute of Fundamental Technological Research Polish Academy of Sciences).

1 Objectives World statistics show that each year the number of newly diagnosed breast cancer cases exceeds 500,000. These puts breast cancer in the first place among other oncological diseases in terms of its incidence in women. Presently, the most frequently is used so-called comprehensive treatment which in-cludes among others neoadjuvant chemotherapy (NAC). The available results of many clinical trials have shown that NAC is a very effective method of treatment which, in many cases results a complete pathological response. However, some studies have also shown that 10–35% of patients do not respond to chemotherapy. The use of ineffective longterm treatment reduces the chances of recovery and survival. Therefore, monitoring, assessment and prediction of the pathological tumor response to NAC should be precisely performed at each stage of therapy. The aim of the study was to assess the breast tumor response to neoadjuvant chemotherapy, after each chemotherapy cycle, using a new ultrasound technique that is able to visualize the relative size of acoustic scatterers - H-scan ultrasound imaging. During successful chemotherapy, the microstructure of tumor is remodeled. This multistep process involves changes in the size of cancer cell nuclei, which is one of the first observable changes in an apoptotic tumor cell and subsequent changes in the stroma. Consequently, the different structures are the main source of ultrasound scattering in the various phases of treatment. Therefore, it can be assumed that changes in the H-scan ultrasound images will be also observed during the treatment process. 2 Materials and Methods Breast ultrasound examinations were performed on 51 tumors. Before qualifying for NAC, all patients were subjected to coreneedle biopsy. Additional, at the end of chemotherapy and after surgery, removed tumors or residual lesions were assessed to determine the number of residual malignant cells (RMC). RMC is a parameter that allows to evaluate the response to NAC. The RMC percentage ranges from 0% to 100%. It is 0% for complete pathological response and 100% for complete non-response. In the presented studies, the RMC cut-off was assumed to be ≤30% for responding tumors. The acquisition of RF ultrasound echoes from tumors was performed using an Ultrasonix SonixTOUCH ultrasound scanner. Data registration was performed before the start of treatment and 7 days after each of four courses of chemotherapy. RF ultrasound data was recorded for each tumor in the following sections: radial, radial + 45◦, anti-radiative and anti-radiative + 45◦. Two parallel convolution filters GH2 (t) and GH8 (t) were applied to isolate the low and high frequency information in echo signal. Using the Hilbert transform, in the next step, the envelope of filtered signals after normalization was determined. The lower frequency backscatter signals were assigned to channel R (red) and higher frequency components to channel B (blue). In order to quantify the images after each chemotherapy course, the percentage of red in the images was analyzed. Data from four tumor cross section were analyzed as a single set. This approach allowed for a more detailed analysis of changes in the tumor microstructure than was possible with only one crosssection. 3 Results Histopathological examination after final NAC and surgery revealed 29 tumors with RMC 0-30% and 22 tumors with RMC ≥ 30%. The mean values of the percentage of red color (PCR) in the H-Scan images collected before NAC and one week after the 1st, 2nd, 3rd and 4th NAC cycle for responders (RMC ≤ 30%) were 38.4%, 42.6%, 56.7%, 79.1%, 81.6%, respectively. In the group of non-responders (RMC ≥30) the same values were 43.2%, 39.7%, 40.1%, 37.8%, 41.6%. Statistical analysis was performed to differentiate the group of responders and non-responders based on the PCR parameter. The statistical significance of differences between the groups responding to NAC and the groups that did not respond was assessed using the p value. Starting from the second cycle of chemotherapy, statistically significant differences were observed ($p < 0.05$). The AUC for the analyzed parameter as a function of chemotherapy cycles was 0.7

(after the 2nd course), 0.83 (after the 3rd course) and 0.86 (after the 4th course). It can be seen that the analyzed parameter was able to distinguish responders and non-responders after the second treatment dose. Very promising results were obtained after the fourth cycle of chemotherapy, when the use of only one quantitative parameter allowed to predict the response to treatment with AUC ≥ 0.86. 4 Discussion The results presented in this study demonstrated the high potential of H-scan images to predict the response of breast cancer patients to chemotherapy early in treatment. The conducted analysis showed that the responders and non-responders showed a significantly different trend in the percentage of red color in the H-scan images at individual stages of treatment. The information available in the literature on changes in tumor microstructure as a result of successful treatment seems to be consistent with the observations in this study. The remodeling of a neoplastic tumor is a multi-stage process, including changes in neoplastic cells and changes in the stroma. These processes are most likely the source of the changes observed in the H-scan images at the beginning of the treatment. Changes in the stroma: fibrosis, elastosis, collagenization, microcalcifica-tion and neovascularization observed in the subsequent treatment phases also change the dominant sources of ultrasound scattering. The presented results confirm the usefulness of quantitative ultrasonic imaging in monitoring results of NAC treatment. In the near future, further studies are planned to verify these preliminary single-center findings.