

# CSBio 2021

## BOOK OF ABSTRACTS

### THE 12<sup>th</sup> INTERNATIONAL CONFERENCE ON COMPUTATIONAL SYSTEMS-BIOLOGY AND BIOINFORMATICS

OCTOBER 14<sup>th</sup>-15<sup>th</sup>, 2021  
ONLINE VIRTUAL CONFERENCE

#### General Chairs

Kitsuchart Pasupa, King Mongkut's Institute of Technology Ladkrabang  
Chee Keong Kwoh, Nanyang Technological University

#### Program Chairs

Sansanee Auephanwiriyaikul, Chiang Mai University  
Nipon Theera-umpon, Chiang Mai University

**CSBio 2021**  
**Book of Abstracts**

The 12th International Conference on Computational Systems-  
Biology and Bioinformatics

October 14 - 15, 2021  
Fully Virtual Conference

**<https://csbio2021.it.kmitl.ac.th>**

# Preface

This volume contains papers presented at the 12th International Conference on Computational Systems-Biology and Bioinformatics (CSBio2021), held during October 14-15, 2021. Due to the outbreak of COVID-19, this year's conference was organized as a fully virtual conference. With the advancement in next-generation molecular technology in generating the high-throughput "omics" data, life science has come to the era of "big data". To extract biological knowledge from the data and translate it into benefits for society (e.g., better medicine and healthcare), novel and advanced computational tools are needed for data analysis. The current trends are to leverage AI and the application of cognitive computing that combines domain knowledge with machine learning. CSBio2021 aims to provide an international forum for researchers, scientists, and industry professionals who are working in these areas to meet at this event to exchange ideas and stimulate research collaborations.

The conference attracted submissions from 11 countries in Europe, Asia, and North America. All submissions were reviewed by the Program Committee, whose members are highly qualified researchers in the conference topic areas. After careful review by the Program Committee, we finally accepted 9 of the 19 submissions. Additional 1 short paper was also selected to present at the conference. In addition to the works presented at the conference, David Ussery (University of Arkansas for Medical Sciences), a well-known researcher in the field of bioinformatics, was our keynote speaker on a timely topic entitled "The three waves of the Covid-19 pandemic hints at a limited genetic repertoire of SARS-CoV-2".

Finally, we would like to thank all program committees, authors, and attendees for participating in the conference as well as the Organizing Committee members for making this conference a memorable event.

**Kitsuchart Pasupa**, King Mongkut's Institute of Technology Ladkrabang, Thailand  
**Chee Keong Kwoh**, Nanyang Technological University, Singapore  
General Co-Chairs

**Sansanee Auephanwiriyaikul**, Chiang Mai University, Thailand  
**Nipon Theera-Umpon**, Chiang Mai University, Thailand  
Program Co-Chairs

# Table of Contents

<b>Preface</b> .....	<b>ii</b>
<b>Table of Contents</b> .....	<b>iii</b>
<b>Program</b> .....	<b>v</b>
Thursday October 14, 2021 .....	<b>v</b>
Friday October 15, 2021 .....	<b>vii</b>
<b>Conference Committees</b> .....	<b>viii</b>
<b>Keynote Speakers</b> .....	<b>x</b>
<b>The Three Waves of the Covid-19 Pandemic Hints at a Limited Genetic Repertoire of SARS-CoV-2</b>	
David Ussery .....	<b>1</b>
<b><u>Section : Disease Classification &amp; Medical and Biomedical Informatics</u></b>	
<b>Analysis of Dynamics and Stability of Hybrid System Models of Gene Regulatory Networks</b>	
Gatis Melkus, Karlis Cerans, Karlis Freivalds, Lelde Lace, Darta Rituma and Juris Viksna .....	<b>2</b>
<b>Pulmonary Artery Visualization for Computed Tomography Angiography Data of Pulmonary Embolism</b>	
Patiwet Wuttisarnwattana, Annop Krasaesin and Poommetee Ketson .....	<b>3</b>
<b>Testing the Effectiveness of CNN and GNN and Exploring the Influence of Different Channels on Decoding Covert Speech from EEG Signals</b>	
Serena Liu and Jonathan Chan .....	<b>4</b>
<b>Deep Learning-based Approach for Corneal Ulcer Screening</b>	
Kasemsit Teeyapan .....	<b>5</b>
<b>Validating Ontology-based Annotations of Biomedical Resources using Zero-shot Learning</b>	
Dimitrios Koutsomitropoulos .....	<b>6</b>
<b>The Effect of PreTraining Thoracic Disease Detection Systems on Large-Scale Chest X-Ray Domain Datasets [Short Paper]</b>	
Shafinul Haque and Jonathan Chan .....	<b>7</b>

**Section : Gene Expression Analysis**

**Inference of Gene Networks from Single Cell Data through Quantified Inductive Logic Programming**

Samuel Buchet, Francesco Carbone, Morgan Magnin, Mickaël Ménager and Olivier Roux ..... 8

**Detection of Markers for Discrete Phenotypes**

Hannes Klarner, Elisa Tonello, Laura Fontanals, Florence Janody, Claudine Chaouiya and Heike Siebert ..... 9

**Section : Modeling and Simulation of Biological Process**

**Optimization Algorithm for Omic Data Subspace Clustering**

Madalina Ciortan and Matthieu Defrance ..... 10

**Spatio-Temporal Evolution of Cellular Automata based Single Nephron Rigid Tubular Model**

Siva Manohar Reddy Kesu and Hariharan Ramasangu ..... 11

# Program

CSBio 2021 is co-located with the 13th International Conference on Information Technology and Electrical Engineering (ICITEE 2021), <https://icitee2021.it.kmitl.ac.th>. All time listed in this abstract book is Indochina Time (ICT / UTC+07:00).

## Thursday October 14, 2021

### Room 1

- 09:00-09:25     **Opening Ceremony**
- 09:30-10:15     **Keynote Session 1 (ICITEE 2021)**  
Prof. Dr. Masanori Sugimoto;  
Hokkaido University, Japan
- 10:20-11:05     **Keynote Session 2 (CSBio 2021)**  
**The Three Waves of the Covid-19 Pandemic Hints at a Limited Genetic Repertoire of SARS-CoV-2**  
Prof. Dr. David Ussery;  
University of Arkansas for Medical Sciences, USA
- 11:05-11:20     Break

### Room 4 - Section: Disease Classification & Medical and Biomedical Informatics

Chair: Dr. Kasemsit Teeyapan;  
Chiang Mai University, Thailand

- 11:20-11:45     **Analysis of Dynamics and Stability of Hybrid System Models of Gene Regulatory Networks**  
Gatis Melkus, Karlis Cerans, Karlis Freivalds, Lelde Lace, Darta Rituma and Juris Viksna
- 11:45-12:10     **Pulmonary Artery Visualization for Computed Tomography Angiography Data of Pulmonary Embolism**  
Patiwet Wuttisarnwattana, Annop Krasaesin and Poommetee Ketson
- 12:10-13:00     Lunch Break

### Room 1

- 13:00-13:45     **Keynote Session 3 (ICITEE 2021)**  
Assoc. Prof. Dr. Marco Anisetti;  
Università degli Studi di Milano, Italy
- 13:45-14:00     Break

#### Room 4 - Section: Gene Expression Analysis

Chair: Dr. Chawan Manaspon;  
Chiang Mai University, Thailand

- 14:25-14:50     **Inference of Gene Networks from Single Cell Data through Quantified Inductive Logic Programming**  
Samuel Buchet, Francesco Carbone, Morgan Magnin, Mickaël Ménager and Olivier Roux
- 14:50-15:15     **Detection of Markers for Discrete Phenotypes**  
Hannes Klärner, Elisa Tonello, Laura Fontanals, Florence Janody, Claudine Chaouiya and Heike Siebert
- 15:15-15:30     Break

#### Room 4 - Section: Modeling and Simulation of Biological Processes

Chair: Asst. Prof. Dr. Navadon Khunlertgit;  
Chiang Mai University, Thailand

- 15:30-15:55     **Optimization Algorithm for Omic Data Subspace Clustering**  
Madalina Ciortan and Matthieu Defrance
- 15:55-16:20     **Spatio-Temporal Evolution of Cellular Automata based Single Nephron Rigid Tubular Model**  
Siva Manohar Reddy Kesu and Hariharan Ramasangu

Friday October 15, 2021

**Room 1**

09:00-09:45 **Keynote Session 4 (ICITEE 2021)**  
Prof. Dr. Basabi Chakraborty;  
Iwate Prefectural University, Japan

09:50-10:35 **Keynote Session 5 (ICITEE 2021)**  
Prof. Dr. Chu Kiong Loo;  
University of Malaya, Malaysia

10:35-10:50 Break

**Room 4 - Section: Disease Classification & Medical and Biomedical Informatics**

Chair: Assoc. Prof. Dr. Sansanee Auephanwiriyaikul;  
Chiang Mai University, Thailand

10:50-11:15 **The Effect of PreTraining Thoracic Disease Detection Systems on Large-Scale Chest X-Ray Domain Datasets**  
Shafinul Haque and Jonathan Chan

11:15-11:40 **Testing the Effectiveness of CNN and GNN and Exploring the Influence of Different Channels on Decoding Covert Speech from EEG Signals**  
Serena Liu and Jonathan Chan

11:40-12:05 **Deep Learning-based Approach for Corneal Ulcer Screening**  
Kasemsit Teeyapan

12:05-12:30 **Validating Ontology-based Annotations of Biomedical Resources using Zero-shot Learning**  
Dimitrios Koutsomitropoulos

12:30-13:15 Lunch Break

**Room 1**

13:15-14:00 **Keynote Session 6 (ICITEE 2021)**  
Assoc. Prof. Dr. Sri Suning Kusumawardani;  
Universitas Gadjah Mada (UGM), Indonesia

14:00-14:15 Break

**Room 1**

16:00-16:15 **Closing Ceremony**



# Conference Committees

## International Advisory Committee

Jonathan H. Chan, King Mongkut's University of Technology Thonburi, Thailand  
Richard F. Wintle, The Hospital for Sick Children, Canada  
David Ussery, University of Arkansas for Medical Sciences, USA  
Sissades Tongsima, National Biobank of Thailand, National Science and Technology  
Development Agency, Thailand

## General Co-Chairs

Kitsuchart Pasupa, King Mongkut's Institute of Technology Ladkrabang, Thailand  
Chee Keong Kwoh, Nanyang Technological University, Singapore

## Program Co-Chairs

Sansanee Auephanwiriyaikul, Chiang Mai University, Thailand  
Nipon Theera-Umpon, Chiang Mai University, Thailand

## Publication Co-Chairs

Navadon Khunlertgit, Chiang Mai University, Thailand  
Chawan Manaspon, Chiang Mai University, Thailand

## Organizing Chairs

Nont Kanungsukkasem, King Mongkut's Institute of Technology Ladkrabang, Thailand  
Vithida Chongsuphajaisiddhi, King Mongkut's University of Technology Thonburi, Thailand

## Financial Chair

Pattanapong Chantamit-O-Pas, King Mongkut's Institute of Technology Ladkrabang,  
Thailand

## Secretary

Sirasit Lochanachit, King Mongkut's Institute of Technology Ladkrabang, Thailand

## Webmaster

Chayanon Sub-r-pa, King Mongkut's Institute of Technology Ladkrabang, Thailand

## **International Program Committee**

Alberto Policriti, Università di Udine, Italy

Anne Siegel, IRISA, France

Daisuke Kihara, Purdue University, USA

Elisabetta De-Maria, Université Côte d'Azur, France

Géraldine Jean, University of Nantes, France

Joëlle Despeyroux, INRIA & CNRS, France

Osamu Maruyama, Kyushu University, Japan

Patiwet Wuttisarnwattana, Chiang Mai University, Thailand

Shunsuke Aoki, Kyushu Institute of Technology, Japan

Sorayya Malek, University of Malaya, Malaysia

Sylvain Soliman, INRIA Saclay Île-De-France, France

Tatsuya Akutsu, Kyoto University, Japan

Thao Dang, CNRS/VERIMAG, France

Yoichi Takenaka, Kansai University, Japan

# Keynote Speakers



Dr. David Ussery,  
University of Arkansas for Medical Sciences, USA

## **Bio:**

Professor David Ussery was born and raised in Springdale, Arkansas. He has been working with bioinformatic analysis of bacterial genomes since the first sequence was published in 1995, and published one of the first text books in the field of Comparative Genomics. His team has published more than 200 papers, which have been cited more than 15,000 times, including three papers with more than a thousand citations. He has been a co-applicant on grants funded totaling more than \$30 million, since 2010. His popular course on Comparative Genomics, taught at The Technical University of Denmark from 1997 - 2015, is now taught in the spring semesters at UAMS; one-week workshops based on this course have been held in North and South America, Europe, Asia, and Africa. Prof. Ussery has collaborative projects with groups in Belgium, Denmark, France, Germany, The Netherlands, Norway, Spain, Sweden, and the UK, as well as in the U.S.

Prior to joining UAMS, Dr. Ussery was the Comparative Genomics Group lead at Oak Ridge National Labs, in Oak Ridge, Tennessee (2013-2016). He led the Comparative Microbial Genomics group at The Technical University of Denmark from 1997 – 2013, where he has successfully supervised more than 20 Ph.D. students in bioinformatics.

Prof. Ussery received a doctorate in Molecular Biology in 1993 from The University of Cincinnati College of Medicine and did a post-doctoral fellowship at Oxford University (1992-1996). He earned his master's degree in biophysical chemistry at the University of New Mexico in Albuquerque. He earned a bachelor's degree in chemistry from William Jewell College (Liberty, Missouri) in 1982, and graduated from Springdale High School (Springdale, Arkansas) in 1978.

# **The Three Waves of the Covid-19 Pandemic Hints at a Limited Genetic Repertoire of SARS-CoV-2**

Dr. David Ussery,

University of Arkansas for Medical Sciences, USA

The genomic diversity of SARS-CoV-2 is the result of a relatively low level of spontaneous mutations that are introduced during viral replication. We can now begin to assess the overall genetic repertoire of this virus, based on the multitude of genome sequences that have been generated for SARS-CoV-2 during the current pandemic. During 2020, a global wave of one variant remained largely unnoticed, possibly because of its members being divided over several sub-lineages (B.1.177 and sub-lineages B.1.177.XX). We collectively call this Janus, and it represents a pivotal change in the dynamics of the pandemic. Janus created a first wave of a dominant variant, after it was eventually replaced by the variant of concern (VoC) Alpha (B.1.1.7), which in turn has now been replaced by Delta (B.1.617.2). These variants, together with the VoCs Beta (B.1.351) and Gamma (P.1) were compared here, and the presence of their conserved mutations in the complete dataset was assessed. Approximately five percent of the 30,000 nucleotides of the SARS-CoV-2 genome can be variable; it seems that the overall genetic repertoire of SARS-CoV-2 is nevertheless relatively limited, with parallel evolution occurring on a large scale. This may be limiting the immunogenic repertoire of the virus.

# Analysis of Dynamics and Stability of Hybrid System Models of Gene Regulatory Networks

Gatis Melkus, gatis.melkus@lumii.lv,<sup>1</sup>  
Karlis Cerans, karlis.cerans@lumii.lv,<sup>1</sup>  
Karlis Freivalds, karlis.freivalds@lumii.lv,<sup>1</sup>  
Lelde Lace, lelde.lace@lumii.lv,<sup>1</sup>  
Darta Rituma, darta.rituma@lumii.lv,<sup>1</sup>  
Juris Viksna, juris.viksna@lumii.lv,<sup>1</sup>

<sup>1</sup>Institute of Mathematics and Computer Science, University of Latvia, Latvia

We present hybrid system based gene regulatory network models for lambda, HK022 and Mu bacteriophages and analysis of dynamics and possible stable behaviours of the modelled networks. Lambda phage model LPH2 is the result of further development of an earlier LPH1 model taking into account more recent biological assumptions about the underlying biological gene regulatory mechanism. HK022 and Mu phage models are new. All three models provide accurate representations of lytic and lysogenic behavioural cycles, and, importantly, allow to conclude that lysis and lysogeny are the only stable behaviours that can occur in the modelled networks. Along with these models we describe also some new analysis techniques for hybrid system model state spaces.

The models also allow to derive switching conditions that irrevocably leads to one of these two stable behaviours (these are consistent with proposed biological models) and also constraints on binding site affinities that are required for biologically feasible lysis and lysogeny processes. One of the derived constraints in LPH2 model is required for lambda lysis cycle feasibility and places conditions on cro protein binding site affinities. This is consistent with the constraint obtained previously for LPH1 model, although parts of state spaces that describe lysis in these models are different. Another constraint on protein cl binding affinities that is required for biologically feasible lysogeny cycle is new (and likely has been overlooked earlier). At the same time dynamics of HK022 model (which, notably, lacks N antitermination protein) turns out to be independent of both these constraints, although the involved genes and binding their sites are very similar. The used HSM system framework also allows to reproduce biologically different lysis-lysogeny switching mechanism that is used by Mu phage. In general the results show that HSM hybrid system framework can be successfully applied to modelling small gene regulatory networks (with up to ~20 genes) and for comprehensive analysis of model state space stability regions.

# Pulmonary Artery Visualization for Computed Tomography Angiography Data of Pulmonary Embolism

Patiwet Wuttisarnwattana, patiwet@eng.cmu.ac.th,<sup>1</sup>

Annop Krasaesin, annop\_krasaesin@cmu.ac.th,<sup>1</sup>

Poommetee Ketson, poommetee\_ke@cmu.ac.th,<sup>1</sup>

<sup>1</sup>Chiang Mai University, Thailand

Pulmonary embolism (PE) is a preventable life-threatening disease that is among the top three most common causes of cardiovascular deaths. Producing an accurate diagnosis can be challenging. Nowadays, computer-aided diagnosis has proven itself to be a useful tool for physicians. However, computers need to recognize the relevant human anatomy as accurately as possible. In case of PE, pulmonary artery is the structure in which the lesion manifests. In this study, we propose a segmentation algorithm that accurately identifies voxels occupied by pulmonary artery in computed tomography angiography (CTA) images. The algorithm consists of three parts: lung mask extraction, pulmonary artery detection, and pulmonary artery connection. The technique involves several morphological operations and thresholding to separate the vessels from the background. The pulmonary artery connection further refined the preliminary vessel contours and improved the accuracy. We evaluated our method with the dataset from a publicly available FUMPE (Ferdowsi University of Mashhad's PE) dataset. The resulting Dice similarity coefficients against the ground truth created by human experts was about  $81\% \pm 1\%$ . The visualizations created by the automatic algorithm was also very similar to that created by human experts. Future works building upon our study may contribute to the better diagnosis of PE.

# Testing the Effectiveness of CNN and GNN and Exploring the Influence of Different Channels on Decoding Covert Speech from EEG Signals

Serena Liu, serenamrtwo.liu@mail.utoronto.ca, <sup>1</sup>  
Jonathan Chan, jonathan@sit.kmutt.ac.th, <sup>2</sup>

<sup>1</sup>University of Toronto, Canada

<sup>2</sup>King Mongkut's University of Technology Thonburi, Thailand

In this paper, the effectiveness of 2 deep learning models was tested and the significance of 62 different electroencephalogram (EEG) channels were explored on covert speech classification tasks using time-series EEG signals. Experiments were done on the classification between “in” and “cooperate” from the ASU dataset and the classification between 11 different prompts from the KaraOne dataset. The types of deep learning models used are the 1D convolutional neural network (CNN) and the graphical neural network (GNN). Overall, the CNN model showed decent performance with an accuracy of around 80% on the classification between “in” and “cooperate”, while the GNN seemed to be unsuitable for time-series data. By examining the accuracy of the CNN model trained on different EEG channels, the prefrontal and frontal regions appeared to be the most relevant to the performance of the model. Although this finding is noticeably different from various previous works, it could provide possible insights into the cortical activities behind covert speech.

# Deep Learning-based Approach for Corneal Ulcer Screening

Kasemsit Teeyapan, kasemsit.t@cmu.ac.th,

Chiang Mai University, Thailand

Corneal ulcer is a common corneal symptom that, upon infection, can lead to destruction of the corneal tissues, resulting in corneal blindness. To ease the corneal ulcer screening process, this paper introduces a deep transfer learning architecture based on various backbone networks to help identify two severity levels of the symptom: Early stage and advanced stage. The total of 15 well-known deep convolutional neural networks are used as the base model. The proposed transfer learning-based architectures are trained, validated, and tested on 426, 143, and 143 fluorescein staining slit-lamp images from the public SUSTech-SYSU dataset. The experimental results show that the best model depends on the choice of evaluation metric, as the performances of most networks are on par with one another. In addition, we report that one ResNet50 model can achieve the best F1 score of 95.04% with the AUC score of 99.12%. This model is further evaluated on an external dataset and its prediction is also explained using Integrated Gradients.



# Validating Ontology-based Annotations of Biomedical Resources using Zero-shot Learning

Dimitrios Koutsomitropoulos, koutsomi@ceid.upatras.gr,

University of Patras, Greece

Authoritative thesauri in the form of web ontologies offer a sound representation of domain knowledge and can act as a reference point for automated semantic tagging. On the other hand, current language models achieve to capture contextualized semantics of text corpora and can be leveraged towards this goal. We present an approach for injecting subject annotations using query term expansion against such ontologies. For the user to have an indication of the use-fulness of these suggestions we further propose an online method for validating the quality of annotations using NLI models such as BART and XLM-R. To circumvent training barriers posed by very large label sets and scarcity of data we rely on zero-shot classification and show that semantic matching can contribute above-average thematic annotations. Also, a web-based validation service can be attractive for human curators vs. the overhead of pretraining large, domain-tailored classification models.

# The Effect of PreTraining Thoracic Disease Detection Systems on Large-Scale Chest X-Ray Domain Datasets

Shafinul Haque, shafinul.haque@mail.utoronto.ca, <sup>1</sup>

Jonathan Chan, jonathan@sit.kmutt.ac.th, <sup>2</sup>

<sup>1</sup>University of Toronto, Canada

<sup>2</sup>King Mongkut's University of Technology Thonburi, Thailand

The COVID-19 pandemic has impacted many countries around the world resulting in the need to develop quick and effective screening methods to ease the burden and overcome the limitations of varying healthcare capacities. Given the nature of the disease, the use of Chest X-ray medical imaging has proven to be very useful which has prompted the exploration of computer-aided diagnosis tools to augment and enhance radiologists. However, recent reports have deemed many of the proposed methods to be impractical for use in real-life applications due to models with poor generalization capabilities, an issue closely related to the quality of current datasets in the CXR domain. Typically, deep convolutional neural network (CNN) based classification systems utilize transfer learning techniques when data is limited. We suggest first training models on publicly available large-scale and CXR specific datasets such as CheXpert and using these pretrained weights when initializing the final model. Compared with a CNN pretrained on the more general ImageNet dataset, pretraining on large-scale domain specific data increased the model's ability to generalize to unseen data.

# Inference of Gene Networks from Single Cell Data through Quantified Inductive Logic Programming

Samuel Buchet, samuel.buchet@ec-nantes.fr, <sup>1</sup>  
Francesco Carbone, francesco.carbone@institutimagine.org, <sup>2</sup>  
Morgan Magnin, morgan.magnin@ec-nantes.fr, <sup>1</sup>  
Mickaël Ménager, mickael.menager@institutimagine.org, <sup>2</sup>  
Olivier Roux, olivier.roux@ec-nantes.fr, <sup>1</sup>

<sup>1</sup>Ecole centrale de Nantes, France

<sup>2</sup>Université de Paris, Imagine Institute, France

Single cell sequencing technologies represent a unique opportunity to appreciate all the heterogeneity of gene expressions within specific biological cell types.

While these data are sparse and contain a lot of noise, it remains possible to perform multiple analysis tasks such as identifying sub cellular types and biological markers. Beyond revealing distinct sub cell populations, single cell gene expressions usually involve complex gene interactions, which may often be interpreted as an underlying gene network.

In this context, logical computational approaches are particularly attractive as they provide models that are easy to interpret and verify. However, the noise is especially important in single cell sequencing data. This may appear as a limit for symbolic methods as they usually fail in addressing the statistical aspect necessary to handle efficiently such noise.

In this work, we propose a computational approach based on symbolic modeling to identify gene connections from single cell RNA sequencing data. Our algorithm, LOLH, is based on Inductive Logic Programming, and intends to rapidly identify potential gene interactions through discrete optimization, while accounting for the noise in the data. By combining symbolic modeling with optimization techniques, we aim to provide an interpretable model that still fits properly on the data.

We apply our method to the unsupervised inference of a gene correlation network from a concrete single cell dataset. We show that the output of our algorithm can be interpreted by using the data itself, and we use additional biological knowledge to validate the approach.

# Detection of Markers for Discrete Phenotypes

Hannes Klarner, hannes.klarner@fu-berlin.de, <sup>1</sup>  
Elisa Tonello, etonello@zedat.fu-berlin.de, <sup>1</sup>  
Laura Fontanals, lcifuentes@zedat.fu-berlin.de, <sup>1</sup>  
Florence Janody, fjanody@ipatimup.pt, <sup>2</sup>  
Claudine Chaouiya, claudine.chaouiya@univ-amu.fr, <sup>3</sup>  
Heike Siebert, siebert@mi.fu-berlin.de, <sup>1</sup>

<sup>1</sup>Freie Universität Berlin, Germany

<sup>2</sup>Instituto Gulbenkian Ciência, Portugal

<sup>3</sup>Aix-Marseille Université, France

Motivation: Capturing the molecular diversity of living cells is not straightforward. One approach is to measure molecular markers that serve as indicators of specific biological conditions or phenotypes. This is particularly relevant in modern medicine to provide precise diagnostics and pinpoint the best treatment for each patient. The challenge is to select a minimal set of markers whose activity patterns are in correspondence with the phenotypes of interest.

Results: This article approaches the marker detection problem in the context of discrete phenotypes which arise, for example, from Boolean models of cellular networks. Mathematically this poses a combinatorial optimization problem with many answers. We propose a solution to this optimization problem that is based on the modelling language answer set programming (ASP). A case study of a death cell receptor network illustrates the methodology.

# Optimization Algorithm for Omic Data Subspace Clustering

Madalina Ciortan, [madalina.ciortan@ulb.be](mailto:madalina.ciortan@ulb.be),<sup>1</sup>  
Matthieu Defrance, [matthieu.dc.defrance@ulb.ac.be](mailto:matthieu.dc.defrance@ulb.ac.be),<sup>1</sup>

<sup>1</sup>Université Libre de Bruxelles, Belgium

Subspace clustering identifies multiple feature subspaces embedded in a dataset together with the underlying sample clusters. When applied to omic data, subspace clustering is a challenging task, as additional problems have to be addressed: the curse of dimensionality, the imperfect data quality and cluster separation, the presence of multiple subspaces representative of divergent views of the dataset, and the lack of consensus on the best clustering method.

First, we propose a computational method (discover) to perform subspace clustering on tabular high dimensional data by maximizing the internal clustering score (i.e. cluster compactness) of feature subspaces. Our algorithm can be used in both unsupervised and semi-supervised settings. Secondly, by applying our method to a large set of omic datasets (i.e. microarray, bulk RNA-seq, scRNA-seq), we show that the subspace corresponding to the provided ground truth annotations is rarely the most compact one, as assumed by the methods maximizing the internal quality of clusters. Our results highlight the difficulty of fully validating subspace clusters (justified by the lack of feature annotations). Tested on identifying the ground-truth subspace, our method compared favorably with competing techniques on all datasets. Finally, we propose a suite of techniques to interpret the clustering results biologically in the absence of annotations. We demonstrate that subspace clustering can provide biologically meaningful sample-wise and feature-wise information, typically missed by traditional methods.

# **Spatio-Temporal Evolution of Cellular Automata based Single Nephron Rigid Tubular Model**

Siva Manohar Reddy Kesu, smreddyyyy@cvsr.ac.in, <sup>1</sup>

Hariharan Ramasangu, rharihar@ieee.org, <sup>2</sup>

<sup>1</sup>M S Ramaiah University of Applied Sciences, India

<sup>2</sup>Relecura Inc, India

Partial differential equations play an important role in mathematical modeling of nephrons. The finite difference solution methods exhibit regular, period doubling and irregular oscillations. In this paper, a single nephron model with transport mechanism and autoregulatory mechanism has been developed using cellular automata framework for a rigid tubule. Cellular automata framework captures the emergent behavior of the system. The importance of cellular automata approach of studying a dynamical system emanates from its ability to capture new behavior not easily shown by numerical analysis. The governing equations of a single nephron model are converted to cellular automata local rules using ultradiscretization. The emergent properties from the local cellular automata rules have been compared with the reported experimental findings. It has been shown that cellular automata framework is a promising approach to model macrolevel behaviors of physiological systems.

# **CSBio 2021**

**THE 12<sup>th</sup> INTERNATIONAL  
CONFERENCE ON  
COMPUTATIONAL  
SYSTEMS-BIOLOGY  
AND BIOINFORMATICS**

<https://csbio2021.it.kmitl.ac.th>