



# Symposium on Machine Learning and Data Modelling in the Biomedical Sciences (MLDMBioMed-2022)

**September 27- 28, 2022**

**Laboratory for Industrial and Applied Mathematics (LIAM)**

**in collaboration with**

**Africa-Canada AI and Data Innovation Consortium (ACADIC)**

**Book of Abstracts**

# Symposium on Machine Learning and Data Modelling in the Biomedical Sciences (Sept 27-28, 2022)

## Invited Speakers



**Prof. Nathaniel Osgood,**

Professor in the Department of Computer Science and Associate Faculty in the Department of Community Health & Epidemiology at the University of Saskatchewan, Canada.

**Title of talk at the symposium:** Machine-Learning enabled dynamic models for service delivery and research for the COVID-19 and Opioid epidemic. [Sep 27, 10:30-11:30 EDT.](#)



**Prof. Jacek Banasiak,**

Professor and DST/NRF SARChI Chair in Mathematical Models and Methods in Biosciences and Bioengineering, University of Pretoria, South Africa; and Professor, University of Technology, Institute of Mathematics, Warsaw, Poland.

**Title of talk at the symposium:** Multiple time scales and long-term dynamics in malaria modelling. [Sep 27, 11:30-12:30 EDT.](#)



**Prof. Wilfred Ndijfon,**

Professor of *Theoretical Biology* at AIMS, and the AIMS Network's Research Director.

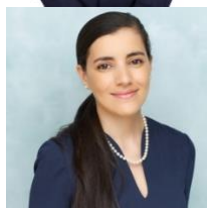
**Title of talk at the symposium:** Data modelling in biomedical sciences: the value of 'first-principles' thinking. [Sep 27, 9:30-10:30 EDT](#)



**Prof. Bruce Mellado,**

Professor at University of the Witwatersrand and iThemba LABS.

**Title of talk at the symposium:** An Intelligent Air Quality Monitoring and Prediction System for Smart Cities. [Sep 28, 11:00-12:00 EDT](#)



**Prof. Bouchra Nasri,**

Assistant professor, Department of Social and Preventive Medicine - Université de Montréal, Canada

**Title of talk at the symposium:** Epidemiological surveillance using Twitter data: some case studies. [Sep 28, 16:00-17:00 EDT](#)



**Dr. Yonas G. Weldelessie,**

Senior Research Fellow at Warwick Medical School, S& Lecturer School of Mathematics & Statistics, The Open University, UK

**Title of talk at the symposium:** Machine learning prediction of early postpartum glucose intolerance in women with gestational diabetes mellitus. [Sep 27, 13:30-14:30](#)

# Machine-Learning Enabled Dynamic Models for Service Delivery and Research for the COVID-19 and Opioid Epidemics

Prof. Nathaniel Osgood,

Director Computational Epidemiology & Public Health Laboratory [CEPHIL] Department of  
Computer Science, University of Saskatchewan

**Abstract:** While dynamic models offer great value in informing public health understanding, planning and response, the ability of public health decision-makers to rely purely upon traditional dynamic models with pre-set assumptions -- no matter how favourably evidenced when built -- is challenged by numerous factors, including diverse unfolding uncertainties. For COVID-19, such uncertainties at varying points in the pandemic include those associated with dynamics of underlying levels of population immunity, variant emergence, changes in public commitment to social distancing and mask use, adherence to and relaxation of measures, and outbreaks in vulnerable communities. Ongoing replanning in this area is associated with prioritizing vaccine rollouts and conducting timely messaging, rolling back and re-instituting measures, initiating surge planning, and other areas. For the opioid epidemic, unfolding uncertainties include those associated with stochastic evolution of street drug availability, varying levels of adulteration, the appearance of new synthetic opioids, and shifts in substance use patterns due to demand- and supply-side factors. This area is associated with ongoing adaptation of cross-sectoral efforts in areas such as prescription monitoring, prevention, harm reduction, treatment, and enforcement strategies, and a need for ongoing reliable understanding of the underlying system state. In both these spheres, data collected by any one indicator often offers a very limited understanding of the underlying situation, and ongoing monitoring, replanning and projection forward can strongly benefit from approaches that continuously integrate into dynamic models data from diverse and complementary unfolding time series, so as to provide a consistent, integrated depiction of the underlying epidemiology and service demand now and in coming weeks and months.

We describe here the design, implementation and day-to-day monitoring and planning use of Machine Learning-enabled compartmental models supported by a purpose-built computational infrastructure that has been widely deployed for COVID-19 and is now seeing use to understand the opioid epidemic. The use of the Sequential Monte Carlo algorithms of Particle Filtering and Particle Markov Chain Monte Carlo allows such models -- for opioids and COVID-19 -- to be regrounded daily by and adapt to new trends within diverse types of incoming data. Each model includes as outputs estimates (via sampling) of key factors for day-to-day monitoring and decision making. Following a brief description of model design, we describe how particle filtering is used to continually reground estimates of dynamic model state, to support probabilistic model projection and to probabilistically evaluate tradeoffs between potential intervention scenarios. We further note aspects of model use and support in practice, including a sketch of the purpose-built distributed computing framework and associated scripting pipeline. This modular infrastructure permits semi-automated model deployment, automatic post-scenario scripting and reporting for a wide variety of clients with differing reporting needs. Finally, we discuss ongoing evolution of these approaches, particularly to enhance performance and support hierarchical modeling.

**Data modelling in biomedical sciences: the value of 'first-principles' thinking**

**Speaker:** Prof. Wilfred Ndifon, the AIMS Network's Research Director.

**Abstract:** There are various ways in which modelling is productively applied to biological systems. Here, I'm concerned with a specific approach that prioritizes deduction by proceeding from "first principles". Using, for illustrative purposes, the process of DNA quantification by means of amplification, I will show that a first principles approach to modelling makes possible a much more accurate quantification of DNA than has thus far been achieved by using purely statistical approaches.

**Multiple time scales and long-term dynamics in malaria modelling**

Prof. Jacek Banasiak,

University of Pretoria, South Africa & Technical University of Łódź, Poland

**Abstract:** Mathematical models of malaria are notoriously difficult due to the presence of (at least) two interacting populations, human hosts and mosquitoes. Fortunately for modellers, these populations evolve at widely different time scales. This offers a way to simplify the model using the tools from the singular perturbation theory. However, to be relevant for the long-term dynamics of the models, the asymptotic results must be uniform in time. In this talk, we present a survey of recent work on these aspects of asymptotic analysis and show how they can be applied to achieve a meaningful reduction in the complexity of some malaria models.

## Machine learning prediction of early postpartum glucose intolerance in women with gestational diabetes mellitus

Authors: Durga Parkhi, Nishanthi Periyathambi, Yonas Ghebremichael-Weldeselassie, Vinod Patel, Nithya Sukumar, Rahul Siddharthan, Leelavati Narlikar, Ponnusamy Saravanan\*

**Speaker:** Dr. Yonas Ghebremichael-Weldeselassie,

Senior Research Fellow at Warwick Medical School, S& Lecturer School of Mathematics & Statistics, The Open University, UK

**Background:** Type 2 diabetes (T2D) and cardiovascular disease (CVD) are serious complications of pregnancies with gestational diabetes mellitus (GDM). Around 50% of the GDM women develop postpartum glucose intolerance (ppGI) within the first decade of pregnancy with 10-fold higher risk of future T2D and 2-fold higher risk of CVD. Therefore, it is crucial to identify women at high risk of ppGI before delivery.

**Aim:** We aim to stratify GDM women during pregnancy according to their risk of developing ppGI within 6-12 weeks postpartum.

**Method:** We use retrospective antenatal data consisting of maternal anthropometric and biochemical variables, collected for 607 women diagnosed with GDM from an NHS Trust hospital in the UK. We build a prediction model using machine learning techniques and sparse logistic regression algorithm with nested cross-validation for hyperparameter tuning and model evaluation. We evaluate the performance of the model using metrics such as area under the receiver operating characteristic (ROC) curve and specificity for predetermined values of sensitivity. We decide the optimal threshold for risk stratification using K-L divergence and information graphs. We perform decision curve analysis to compare the standardized net benefit obtained from implementing our prediction model in comparison with universal screening of GDM women for ppGI.

**Results:** 92 of women developed ppGI. Antenatal fasting plasma glucose (A-FPG) at the time of Oral Glucose Tolerance Test (OGTT) and antenatal HbA1c (A-HbA1c) immediately after the diagnosis of GDM are the only two variables selected by our model in >80% of the iterations. The model has an area under the ROC curve of 0.72. We propose a rule-in diagnostic test with 92% specificity at a probability threshold of 0.381 and a rule-out diagnostic test with 92% sensitivity at a probability threshold of 0.140.

**Conclusion:** We propose a ‘rule-in and rule-out’ approach to predict ppGI. This will allow risk stratification of GDM women before delivery and a targeted, personalized approach. We believe this model is an easy practical tool for healthcare professionals and policy makers in both high and low resource settings by varying the threshold at which risk reduction strategies are implemented in GDM women for reducing cardiometabolic outcomes.

# Tracing Unemployment Rate of South Africa during the COVID-19 Pandemic Using Twitter Data

Zahra Movahedi Nia<sup>1</sup>, Ali Asgary<sup>2</sup>, Nicola Bragazzi<sup>1</sup>, Bruce Mellado<sup>3</sup>, James Orbinski<sup>4</sup>, Jianhong Wu<sup>1</sup>, Jude Kong<sup>1,5</sup>

<sup>1</sup>Africa-Canada Artificial Intelligence and Data Innovation Consortium (ACADIC), Laboratory for Industrial and Applied Mathematics, York University, Canada

<sup>2</sup>Africa-Canada Artificial Intelligence and Data Innovation Consortium (ACADIC), the Advanced Disaster, Emergency and Rapid Response Program, York University, Canada

<sup>3</sup>Africa-Canada Artificial Intelligence and Data Innovation Consortium (ACADIC), School of Physics, Institute for Collider Particle Physics, University of the Witwatersrand, Johannesburg, South Africa

<sup>4</sup>Africa-Canada Artificial Intelligence and Data Innovation Consortium (ACADIC), the Dahdaleh Institute for Global Health Research, York University, Canada

<sup>5</sup>[jdkong@york.ca](mailto:jdkong@york.ca)

## Abstract

Global economy has been hardly hit by the COVID-19 pandemic. Many countries are experiencing a severe and destructive recession. Significant number of firms and businesses have gone bankrupt or been scaled down, and many individuals have lost their jobs. The main goal of this study is to support policy- and decision-makers with additional and real-time information about the labor market flow during a prolonged pandemic. We use social media data, particularly, Twitter to trace and nowcast the unemployment rate of South Africa during the COVID-19 pandemic. We use the quantity and quality of the tweets to estimate the unemployment rate in real-time, since it usually takes months for formal unemployment data to be published. Moreover, we find the missing unemployment rates in cases where census measurements are incomplete. Due to on and off rounds of lockdowns, unemployment rate has changed rapidly during the COVID-19 pandemic and quarterly estimations are not capable of capturing the fluctuations. Therefore, we first used Google mobility index to interpolate the quarterly unemployment rate and find the monthly values. Next, we created a dataset of unemployment related tweets in South Africa using certain keywords. Principal Component Regression (PCR) was applied to nowcast the unemployment rate using the gathered tweets and the sentiment scores of them. Numerical results indicate that the volume of the tweets has a positive correlation, and the sentiment of the tweets has a negative correlation with the unemployment rate during and before the COVID-19 pandemic. Moreover, the nowcasted unemployment rate using PCR has a low Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric MAPE (SMAPE), and a high R<sup>2</sup>-score.

**Keywords:** sentiment analysis, social media, Twitter data, Google mobility index, unemployment rate, labor market, COVID-19, South Africa.

## Pharmaceutical and Non-Pharmaceutical Interventions for Controlling the COVID-19 Pandemic

S. Farhang-Sardroodi<sup>1\*</sup>, I. R. Moyles<sup>2</sup>, J. M. Heffernan<sup>2,3</sup>  
<sup>1</sup>Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, Canada  
<sup>2</sup>Centre for Disease Modelling (CDM), Mathematics Statistics, York University, Toronto, Canada  
<sup>3</sup>Modelling Infection and Immunity Lab, Mathematics Statistics, York University, Toronto, Canada

\*Suzan.Farhangsardroodi@umanitoba.ca

**Abstract:** Disease prevalence can be affected by pharmaceutical (drug) and non-pharmaceutical interventions (such as social distancing, mask-wearing, and contact tracing). Understanding the interaction between disease dynamics and human behavior on one side and the pharmacodynamics on the other are significant factors to control infectious. In this presentation, we propose a 5-compartment epidemiological model for studying how infectious diseases evolve for people with different levels of immunity, natural immunity, and vaccine-induced immunity. Since certain people are confirmed to be less protected by current vaccine regimens, two different cohorts have been assumed for vaccinated people. We increased the complexity of the underlying mechanistic model by adding different waning immunity parameters to capture more of the system dynamics. The model was applied to the record data from Ontario, Canada, and the parameter fittings led to an excellent agreement beyond the time window from March 10, 2020, up to November 30, 2021.



# Identification of candidate antibody targets for disease diagnosis using high-throughput data and novel analytical pipelines

André Filipe Afonso de Sousa Fonseca (0000-0001-8249-0354)<sup>1</sup>

<sup>1</sup> PhD in Biotechnological Science  
a49406@ualg.pt

**Abstract:** In the previous years, we developed analytical pipelines for antibody data with the objective of identifying antimalarial vaccine candidates. We have now extended these pipelines to deal with high dimensional datasets containing thousands of antibodies under analysis. Several concerns have been addressed to increase the robustness of our analysis in these high dimensional settings. Multicollinearity and correction for multiple testing have been studied to reduce the chance of finding false positives. Additionally, alternative approaches based on Random Forest, Gradient Boosting and Neural Networks have been implemented to make our analysis sturdier. To demonstrate the utility of these techniques, we have compared their performances with the ones generated from traditional statistical methods. This comparison allowed us not only to understand which approach performed the best, but also if these machine learning techniques were performing better than traditional ones, and if so, what was the increment in performance.

To illustrate our new pipelines, we have analysed with two challenging datasets: 1) published data related to antibody responses against 2320 *P.falciparum*-specific peptides measured in 186 Malian individuals<sup>1</sup>; 2) published data of antibody responses against 3054 peptides derived from Epstein-Barr virus in 92 individuals with Myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) and 50 healthy individuals<sup>2</sup>. According to our results, we were able to construct classifiers that reached areas under the curve (AUC) of the Receiving Operating Characteristic (ROC) curve of at least 0.90 in both datasets.

In summary, we have improved our pipelines by making them more accurate and allowing them to handle larger dataset more efficiently and precisely. Moreover, by extending our pipeline to contexts other than malaria while keeping similar performances, we demonstrate that they can be employed to any infection's disease where antibody data is an important research component.

## Acknowledgments

I am grateful for my PhD fellowship granted by FCT– Fundação para a Ciência e Tecnologia (ref. SFRH/BD/147629/2019). I am thankful for Przemysław Biecek for letting me work alongside his group in Faculty of Mathematics and Information Science, Warsaw University of Technology, located in Warsaw, Poland

## Bibliography

1. Valletta JJ, Recker M. Identification of immune signatures predictive of clinical protection from malaria. *PLOS Computational Biology*. 2017;13(10):e1005812. doi:10.1371/journal.pcbi.1005812

2. Loebel M, Eckey M, Sotzny F, et al. Serological profiling of the EBV immune response in Chronic Fatigue Syndrome using a peptide microarray. *PLoS ONE*. 2017;12(6). doi:10.1371/journal.pone.0179124

**Title: Modeling high-dimensional interaction problems with the pliable lasso**

**Author:** Theophilus Quachie Asenso, Manuela Zucknick

**Affiliation:** Oslo Centre for Biostatistics and Epidemiology (OCBE), University of Oslo

**Abstract:** Modelling interactions in high-dimensional data is a notoriously difficult problem. Analyzing high-dimensional data with conventional tools is very challenging. One of the reasons is that most existing models cannot easily handle cases with high-dimensional data and many interaction effects among the covariates, and they often make strong assumptions, e.g. strong hierarchy between main and interaction effects. In this work we apply the pliable lasso penalty to estimate interaction effects and extend the existing linear pliable lasso model to other models. We also demonstrate how the pliable lasso can be used to estimate interactions.

We present two extensions of pliable lasso, including the **MADMM algorithm for multi-response data**. We further apply the **MADMM** algorithm to **cancer drug sensitivity screening** data for personalized cancer therapy.

## Synchronization transitions caused by time-varying coupling functions

**Speaker:** Zeray Hagos Gebrezabher , University of Sao Paulo, Brazil.

Interacting dynamical systems are widespread in nature. The influence that one such system exerts on another is described by a coupling function; and the coupling functions extracted from the time-series of interacting dynamical systems are often found to be time-varying. Although much effort has been devoted to the analysis of coupling functions, the influence of time-variability on the associated dynamics remains largely unexplored. Motivated especially by coupling functions in biology, including the cardio-respiratory and neural delta-alpha coupling functions, this paper offers a contribution to the understanding of effects due to time-varying interactions. Through both numerics and mathematically rigorous theoretical consideration, we show that for time- variable coupling functions with time-independent net coupling strength, transitions into and out of phase- synchronization can occur, even though the frozen coupling functions determine phase-synchronization solely by virtue of their net coupling strength. Thus the information about interactions provided by the shape of coupling functions plays a greater role in determining behaviour when these coupling functions are time-variable.

**Title: Machine learning method to learn the stochastic dynamical systems and its application in polymer dynamics**

**Speaker:** Dr. Xiaoli Chen, University of Singapore, Singapore

**Abstract:**

In this talk, I will introduce the machine learning method to learn the stochastic problem. First, I will introduce how to combine physics informed neural network (PINN) method and the sample observation data to learn the stochastic differential equation driven by Brown and Levy noise. Then, I will introduce using stochastic OnsagerNet to learn closure dynamical systems. Lastly, I will apply our method in polymer dynamics.

## **The estrogen paradox in breast cancer treatment. Mathematical modelling and optimal control.**

**Speaker:** Prof. R. Ouifki,  
North-West University, Mahikeng Campus, South Africa

### **Abstract:**

**Background:** Estrogen is known to stimulate the growth of breast cancer, but is also effective in treating the disease. This is referred to as the “estrogen paradox”. Furthermore, short-term treatment with estrogen can successfully eliminate breast cancer, whereas long-term treatment can cause cancer recurrence. Studies highlighted clinical correlations between estrogen and the protein p53 which plays a pivotal role in breast cancer suppression.

**Aim:** We sought to investigate how the interplay between estrogen and p53 impacts the dynamics of breast cancer, and further explore if this could be a plausible explanation for the estrogen paradox and the paradoxical tumor recurrence that results from prolonged treatment with estrogen. Finally, we formulate and analyse an optimal control problem to optimize breast cancer treatment using hormonal therapy with estrogen and Gene therapy using p53.

**Methods:** For this, we propose a novel ODE based mathematical model that accounts for dormant and active cancer cells, along with the estrogen hormone and the p53 protein. We analyze the model’s global stability behavior using the Poincaré-Bendixson theorem and results from differential inequalities. We also perform a bifurcation analysis and carry out numerical simulations that elucidate the roles of estrogen and p53 in the estrogen paradox and its long-term estrogen paradoxical effect.

**Results:** The mathematical and numerical analyses suggest that the apparent paradoxical role of estrogen could be the result of an interplay between estrogen and p53. More importantly the mathematical findings provide explicit conditions under which the paradoxical effect of long-term treatment may be prevented.

**Key words:** Breast cancer - Estrogen paradox - Mathematical modelling - Ordinary differential equations - Global Stability - Global Bifurcation - Optimal control.

## **Ensemble of Deep Learning Models for Early Detection of Monkeypox using clinical symptoms and digital images**

Otasowie Owolafe (Ph.D),

Department of Cyber Security, Federal University of Technology, Akure, Nigeria [oiyare@futa.edu.ng](mailto:oiyare@futa.edu.ng)

### **ABSTRACT**

The continuous and rapid spread of the monkeypox (MPX) virus is become worrisome to health communities and the global world. From the first human recorded infection in 1970 in Congo, the disease has spread to recognizable number of countries around the world affecting more than 16,000 individuals. Of these numbers, Nigeria is having a total of 1,216 cases spread across the different parts of the country. This means that at least 75 countries have been affected so far and the World Health Organisation (WHO) has declared MPX an emergency so that attention will be focused on the outbreak to spur action and a coordinated international response. Past researches have proposed different approaches in inhibiting the spread of the disease as a means of aiding health practitioners in the easy identification of the virus. Most of these literatures made use of small dataset and this significantly reduces the generalizability of the model. However, to be successful in this regard it is imperative to be equipped with a solution that can perform early and automatic detection of MPX that can learn from experience without explicitly programmed to do so. Therefore, this research is aimed at proffering solutions to these problems by developing an ensemble deep learning models for detecting Monkeypox disease in an environment characterized by python and its libraries. The developed model will be evaluated on the basis of accuracy, specificity and recall using dataset that encompasses all variants of the pox family. It is expected that the model aid in successfully curbing the spread of the disease by producing a distinctive clinical presentations for clinicians to enable them easily distinguish monkeypox from other rashes.

## Identification of sex-specific targets of renal ischemia-reperfusion injury (IRI) using artificial intelligence.

Stéphane Nemours<sup>1</sup>, Anna Meseguer<sup>2,3,4</sup>

1. Molecular Oncology Group. Biodonostia Health Research Institute. Paseo Dr. Begiristain, s/n, 20014 San Sebastián, Spain.
2. Renal Physiopathology Group, Vall d'Hebron Research Institute, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain.
3. Departament de Bioquímica i Biologia Molecular, Unitat de Bioquímica de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain.
4. Red de Investigación Renal (REDINREN), Instituto Carlos III-FEDER, Madrid, Spain.

**Abstract:** Men are more prone to acute kidney injury (AKI) and chronic kidney disease (CKD) than women. Renal ischemia-reperfusion injury (IRI), which is faced in many clinical situations, is a major cause of AKI leading to injury and death of proximal tubule epithelial cells (PTEC). It is accepted that regeneration by surviving PTEC is the predominant mechanism of repair/regeneration after ischemic tubular injury in the adult mammalian kidney. In this study, based on recently published transcriptomic data of male and female ischemic kidney biopsies at different time points (1), sex-specific prediction models of IRI were generated using Anaxomics Biotech S.L. Therapeutic Performance Mapping System (TPMS) approach. Key proteins and their mechanism of action were identified. Among other targets, the transcription factor STAT3, regulated by androgens, was determined as a central proteins involved in different processes characterizing IRI such as oxidative stress, immune responses and regeneration of epithelial cells. To validate some findings, an *in vitro* renal IRI cellular model was established. In particular, it was revealed that STAT3 is regulated at the phosphorylation level during renal IRI through serine 727 and tyrosine 705 residues. Overall, this study constitutes an extensive characterization of androgen's targets regulation upon renal IRI. Data provided shall help to better understand the sex differences in kidney diseases at the molecular level.

This work was supported in part by grants from Ministerio de Economía y Competitividad (SAF2014-59945-R and SAF2017-89989-R to A. Meseguer) and Red de Investigación Renal REDinREN (12/0021/0013 to A. Meseguer).

### Reference

1. Nemours, S., Castro, L., Ribatallada-Soriano, D. et al. Temporal and sex-dependent gene expression patterns in a renal ischemia-reperfusion injury and recovery pig model. *Sci Rep* **12**, 6926 (2022). <https://doi.org/10.1038/s41598-022-10352-3>

Title: **An Intelligent Air Quality Monitoring and Prediction System for Smart Cities**

Speaker: **Prof. Bruce Mellado,**

Institute for Collider Particle Physics & ACADIC, University of the Witwatersrand and iThemba LABS



## Classification Analysis of Some Cancer Types by Using Machine Learning

**Speaker:** Scott Ulrich Jemea Ebolo,

African Institute for Mathematical Sciences (AIMS) Cameroon, 608 Crystal garden, Limbe, Cameroon., and AI and Industrial Robotics Engineer, National Higher Polytechnic School of Douala, PK17 Logbessou, Douala, Cameroon.

### **Abstract:**

Cancer is a disease caused by changes in deoxyribonucleic acid (DNA), which attacks cells in the body, causing them to grow uncontrollably and spread to other parts of the body. Cancers can be deadly. The fact that it can develop anywhere in the body gives rise to many types of cancer. In order to reduce the mortality rate from cancer, many diagnostic methods have been developed as the appropriate treatment option and are highly dependent on the type of cancer. In this work, we address this issue of classification of some cancer types by using supervised learning methods to classify prostate cancer, lymphoma, leukaemia and small round blue cell tumour (SRBCT). To be more specific, we used five models: support vector machine, decision tree, random forest, K-nearest neighbours and artificial neural network. Each cancer dataset was trained using each of the machine learning methods in Google Colab GPU. The test samples were classified for each cancer type and the performances of the five models were compared in terms of their percentage according to some metrics. To reduce the dimension of the data, we have incorporated a new approach which involves performing PCA on our dataset. This new approach led to the discovery that the KNN method was the best according to our dataset, with 90% accuracy for prostate and 100% for others.

**Keywords:** Classification, cancer, lymphoma, leukaemia, SRBCT, prostate, machine learning techniques, PCA.

## Modelling and analysis for infectious disease epidemics in South Africa

### Modeling Insulin Sensitivity

Nkam Ntambo Roger<sup>1</sup>, Ziad Taib<sup>2</sup>

<sup>1</sup> African Institute for Mathematical Sciences (AIMS)-Cameroon Alumnus

<sup>2</sup> Professor, Chalmers University of Technology, Sweden

**Speaker:** Nkam Ntambo Roger

**Abstract:** Glucose is very essential in the body for energy but excess blood glucose is a sign of a disease called diabetes mellitus caused by the inability of the body to either produce sufficient insulin or use the available insulin to utilize glucose. In this study, we investigated how the hormone responsible for supplying glucose to muscle tissues maximizes glucose usage by these tissues. This was done using an Intra Venous Glucose Tolerance Test dataset that contained measurements for both type-2 diabetes and healthy subjects together with the Bergman's deterministic minimal model. The modeling was based on a nonlinear mixed effects model performed using R-Studio to define and estimate insulin sensitivity. The insulin sensitivity estimates obtained for both type-2 diabetes and healthy subjects were used to study the relationship that exists between insulin sensitivity and physiological parameters like Body Mass Index and gender. Among the results obtained were the following: the main result of the essay is a model based definition of insulin sensitivity and its implementation on empirical data. Based on the derived estimates of insulin sensitivity, we found that, healthy subjects were more sensitive to insulin than type-2 diabetes subjects and more peculiar was the fact that females were more sensitive to insulin than males. As the Body Mass Index increases the subjects became less sensitive to insulin, thus requiring large amounts of insulin to lower blood glucose level. It is therefore of importance to know one's Body Mass Index because obesity is a risk factor for type-2 diabetes. More importantly, we demonstrated that insulin sensitivity can be used as an endpoint in clinical trials and illustrated how sample sizes of such trials can be calculated.

## **Numerical solvability for a class of chaotic systems of differential equations: the Lorenz model, using both classical and fractional calculus.**

**Speaker:** Ntembue Tshiakuisha Chadrac, AIMS-Ghana

### **Abstract:**

Chaos theory is a very important field of application of fractional calculus. During the last few years, the dynamical systems of fractional order have attracted the attention of several researchers. The fact that the system of differential equations with derivative of fractional order such as the Lorenz system exhibits a chaotic behavior, this pushed the researchers to propose some differential and integral operators in order to predict chaotic behavior. In this study, we use the classical method of finite differences and the Caputo-Fabrizio fractional derivative to approach the solutions of Lorenz model, we also present some numerical simulations for each approach, we study the results of two approaches, we compare them, and finally we have observed that, there are some similarity between the two types of models in both classical and fractional calculus. The graphics present some similar behavior and the advantage with the fractional calculus is that we can vary the derivative order (between 0 and 1), that surely will impact the observed chaos. Whereas in the classical calculus the order is fixed at 1. The fractional calculus also includes the classical calculus one when the derivative is 1.

### **Objective:**

Due to the mathematical complexity contained in the chaotic Lorenz model, this project aims to solve the chaotic Lorenz model by classical calculus and fractional calculus. This project also presents some numerical simulations, compare the two methods and finally give some recommendations for improvement.

### **Conclusion**

The elaboration of this project was an opportunity for us to manipulate the chaotic dynamics system of Lorenz, the differential and integral calculus and especially to introduce ourselves to the fractional calculus which until yesterday was for us only information. As we have specified in the introduction that the chaotic system of Lorenz is a dynamical system of fractional order which presents certain characteristics such as nonlinear, aperiodic, deterministic, sensitivity to initial condition, etc. The fractional calculus which was once proposed for differential and integral calculus of non integer order has now become an indispensable tool in the modeling of chaotic dynamic systems.

In this essay, we have presented the Lorenz model, its chaotic behavior, we have used the finite difference method to solve the chaotic model of Lorenz for the classical calculus and for fractional calculus, we have used the Caputo-Fabrizio fractional derivative to solve the chaotic Lorenz system, and finally we have compared these two methods and we comment the results.

# The Impact of City Mobility on Propagation of Infectious Diseases: Mathematical Modelling Approach

Asrat Mekonnen Belachew <sup>\*1,2</sup> and Tiago Pereira<sup>1</sup>

<sup>1</sup>Institute of Mathematical and Computer Sciences, University of São Paulo, São Carlos, Brazil

<sup>2</sup> Department of Mathematics, Ambo University, Ambo, Ethiopia

**Speaker:** Asrat Mekonnen Belachew

## Abstract

Infectious diseases cause morbidity and mortality to an individual and collapse the social, economic, and political systems of the whole world collectively. Mathematical models are fundamental tools and provide a comprehensive understanding of how infectious diseases spread, and designing the control strategy to mitigate infectious diseases from the host population. Modeling the spread of infectious diseases using a compartmental model of inhomogeneous populations is good in terms of complexity. However, in the real world, there is a situation that accounts for heterogeneity, such as ages, locations, and contact patterns of the population which are ignored in a homogeneous setting. In this work, we study how classical an SEIR-infectious disease spreading of the compartmental model can be extended by incorporating the mobility of population between heterogeneous cities during an outbreak of infectious disease. We have developed an SEIR- multi-cities epidemic spreading model using a system of  $4k$  ordinary differential equations to describe the disease transmission dynamics in  $k$ -cities during the day and night. We have verified that the model is epidemiologically (i.e., variables have biological interpretation) and mathematically (i.e., a unique bounded solution exists all the time) well-posed. We computed the next-generation matrix (NGM) for the model and the basic reproduction number  $R_0$  for SEIR-epidemic spreading model with cities mobility. The disease-free equilibrium point of the model is locally asymptotically stable if  $R_0 < 1$  and, unstable if  $R_0 > 1$ . Using the Perron-Frobenius theorem for non-negative matrices and eigenvalue perturbation theorem, we showed that sending a fraction of the population between cities decreases the reproduction number of diseases in interconnected cities. As a result, disease transmission decreases in the population.

**Keywords:** SEIR-model, Mathematical model, City Mobility, Epidemic spreading, Eigenvalue perturbation.

## Identifying COVID-19 Vaccine Hesitancy Hotspots in Nigeria: Analysis of Social Media Posts Research Problem

**Speaker:** Dr. Blessing Ogbuokiri; Department of Mathematics and Statistics, York University, Canada

One of the major challenges faced by health policymakers in the fight against community-based infectious diseases, such as COVID-19, Malaria, Monkeypox, and Marburg, is vaccine hesitancy. In Nigeria, Twitter is one of the social media platforms used to promote anti-vaccination posts. Anti-vaccination posts or reactions on Twitter can lead to a compromise of community confidence or lack of willingness in taking the vaccine during an outbreak.

**Method:** In this research, we collected 10,000 vaccine-related geotagged Twitter posts in Nigeria, from December 2020 to February 2022, to identify hotspots by clustering tweet sentiments. We used the Natural Language Processing pre-trained model known as VADER to classify the tweets into three sentiment classes (positive, negative, and neutral).

**Result:** The outputs were validated using machine learning classification algorithms, including, Naïve Bayes with an accuracy of 66%, Logistic Regression (71%), Support Vector Machines (65%), Decision Tree (61%), and K-Nearest Neighbour (56%). The average Area under the Curve score of 78%, 85%, 83%, 67%, and 63%, respectively, was used to evaluate the quality of the multi-classification outputs. The classified sentiments were visualised on the Nigerian map using *ArcGIS Online*. The point-based location technique was used to calculate the hotspots on the map. Green, red, and grey were used to identify the dominance of positive, negative, and neutral sentiments.

**Limitations:** The dataset used in this research only reflects the opinion of Twitter users whose geolocation was Nigeria from December 2020 to February 2022. Nigeria has a population of about 210 million people with only 32 million active social media users as at January, 2022. Therefore, this research does not, at large, provide the opinion of the people of Nigeria regarding COVID-19 vaccines. However, this research only classified tweets according to their sentiments and clustered the same to identify hotspots.

It is also relevant to state that most Natural Language Toolkit for sentiment analysis techniques do not have the capacity to properly label figurative language, such as sarcasm. However, since the approach we used was able to label and score a large amount of the tweets in our dataset and was verified with the manual labelling of randomly selected (10%) of the tweets, in addition to the 71% accuracy achieved with the Logistic Regression classifier, we assume it was able to deal with the noise generated by this obvious challenge.

**Conclusion:** The outcome of this research shows that social media data can be used to complement existing data in identifying hotspots during an outbreak. It can also be used to inform health policy in managing vaccine hesitancy.

**Keywords:** COVID-19, vaccine hesitancy, hotspot analysis, social media, natural language processing

**Title of talk: Epidemiological surveillance using Twitter data: some case studies**

**Speaker:** Prof. Bouchra Nasri, Department of Social and Preventive Medicine - Université de Montréal, Canada

**Abstract:**