AIMS | African Institute for Mathematical Sciences NEXT EINSTEIN INITIATIVE

QUANTUM LEAP AFRICA

Washington University in St. Louis

DeepMind

UR

AIMS DTP - DATA SCIENCE STUDENTS' DETAILS - COHORT 2

S/N	STUDENT'S NAME	Gender	Nationality	SUPERVISORS	SUPERVISORS' AFFILIATION	GRADUATING INSTITUTION	STUDENT'S RESEARCH TOPIC	ABSTRACT
1	P´eguy Kem-Meka Tiotsop Kadzue	M	Cameroon	Dr.Ismail Yunus Akhalwaya	IBM Research, South Africa	University of Ngaoundere, Cameroon	Quantum Topological Data Analysis for Graph Neural Networks and Applications	Graph neural networks (GNNs) are powerful neural network architectures employed in graph learning tasks. They have already proven tremendously useful in solving graph and node classification problems. However, GNNs cannot capture certain topological structures which are highly relevant for applications that benefit from more accurate connectivity-respecting modeling. This is the case in many applications including image analysis and bioinformatics. This limitation has recently been addressed by the use of persistent homology, a method from the field of Topological Data Analysis, for capturing the topological information of graphs. Even more recently and impressively, the Topological Graph Layer (TOGL) method has introduced a way to combine GNNs with a differentiable version of persistent homology. However, persistent homology is computationally expensive and as a result, neither higher-order TOGL nor TOGL's integration with higher-order GNNs has been considered. In this thesis, we propose addressing these computational limitations and performing these generalizations. Possible computational solutions include using classical parallel computation and quantum topological data analysis run on quantum computers. In addition, this thesis aims to study use cases, possibly including brain, genome, and climate data that would benefit from our new proposed high-order, computationally faster graph approach while not being affected by new additional constraints.
				Dr.Colleen M. Farrelly	Lead Data Scientist at Turing Labs Inc			
				Dr.Yae Ulrich Gaba	QLA -AIMS			
				Yang-Hui He	Oxford University/LIMS			
				Marcos Crichigno	Private supervisor			
				Isaac Nape	University of the Witwatersrand			
				Henry ADAMS	Colorado State University,USA			
2	Brenda Anague	F	Cameroon	Dr.Issa Karambal	QLA -AIMS	University of Rwanda.	Physics-Informed Learning Machine with Application in Climate Science	A large number of natural phenomena are described by a class of equations, called partial differential equations (PDEs), whose analytical resolution might be remarkably complex. The first methods aiming to solve PDEs were developed in the field of numerical analysis. However, these methods, also known as traditional methods, face some limits. For instance, when applying to high-dimensional problems, they fail to approximate the solution accurately, or sometimes, they deliver results that do not respect the physical law described by the differential equation. In addition, many real physical problems following are described with higher dimensions that can bring traditional methods to their limits resulting in poor approximations of solutions. Nowadays, the great potential of machine learning concepts as well as the remarkable increase in computational power in the last years enables scientists to approach PDEs through concepts from the field of machine learning. The socalled physics-informed neural networks are algorithms both making use of machine learning and respecting the physical law that describes the problem the network is trained to approximate the solution respecting the physical law, by embedding the differential equation into the algorithm. Physics-informed neural networks can be adapted to different types of partial differential equations and hence, they can be applied to problems in a wide range of domains including also climate science.
				Dr. Assionvi Hove Kouevi	QLA -AIMS			
				Dr. Stephen Edward Moore	University of Cape coast, Ghana			
				Dr. Bamdad Hosseini	University of Washington, USA			
				Dr. Paterne Gahungu	AIMS			
3	Ifeoma Veronica Nwabufo	F	Senegal	Dr David Richards	University of Exeter, UK	University of Cape Coast, Ghana	How the human embryo develops: combining mathematical modelling and data science	There has been growing interest in the need to understand how the human embryo develops, not least because it could lead to improved ways of selecting embryos in assisted conception (such as IVF). Due to ethical and logistical reasons, it has traditionally been difficult to study human embryos. However, the recent emergence by our collaborators of a stem-cell model (called the blastoid) that mimics the earliest stages of embryogenesis promises to revolutionise this field, providing vast amounts of data that can inform and guide mathematical modelling. This PhD will, for the first time, utilise this blastoid data to develop a model of the first week of development. This will involve using a combination of data science approaches (including deep learning), mathematical modelling, computer simulation and image analysis. Long term, this work aims to improve how embryos are selected in the IVF clinic, providing improved methods to score early embryos for those most likely to implant and develop successfully.
				Prof. Nick Monk	AIMS Ghana			
4	Esaie Dufitimana	M	Rwanda	Prof. dr. Ate Poorthuis	Katholieke Universiteit Leuven, Belgium	University of Rwanda	Modelling urban socio-economic development and its implication on public health in Kigali city, Rwanda through big data analysis and visualization	Urban poor have increased in many cities, especially in the Global South. This makes effective urban governance even more urgent to help achieve socio-economic development, reduce inequality and build just cities with improved public health. Detailed information is necessary to support this through policy and decision making. However, it is missing. Like other Global South countries, Rwanda is facing lack of necessary information, making it hard to truly understand issues of socio-economic inequality and equity, and their implications on public health in Kigali, the capital city of Rwanda. This research leverages the potential of big data to map and analyse urban socio-economic development and its implication on public health in Kigali at the granular spatial scale of the neighbourhood by combining satellite images, mobile phone data, and data related to the built environment and access to services. Thus, through effective geospatial visualization, the research will translate academic analyses into practical insights to support urban governance, urban management, urban public health, decision-making and defining the priority areas for urban socio-economic transformation to achieve equitable, spatially just cities and improved public health.
				Dr. Ernest UWAYEZU	University of Rwanda			
				Prof. dr. Jean Pierre BIZIMANA	University of Rwanda			
				Dr. Paterne GAHUNGU	AIMS			
5	Bernadin Tamo Amougou	M	Cameroon	Marcelo Pereyra	Heriot Watt University	Heriot Watt University	On the Connection Between Neural Networks and Reproducing Kernel Hilbert Spaces with a View Towards PDE Solvers	Artificial neural networks have achieved state-of-the-art performance across a wide range of machine learning tasks. However, despite their great empirical success they are not understood very well: why do they actually work and why and when do they fail? Through the lens of kernel methods, and a recent mathematical tool, the neural tangent kernel, we aim to understand the dynamics of the training of deep neural networks. This will allow us to characterise the training and the generalisation properties, which are central to describing and leveraging recent advances in the theory of high-dimensional statistical estimation. Physics-informed neural networks (PINNs) are trained to solve supervised learning tasks while respecting any given laws of physics described by general nonlinear partial differential equations. PINNs provide an interesting set-up to investigate the gap between deep learning and kernel learning. New insights in these areas will pave the way towards developing the connection between neural networks and reproducing kernel Hilbert spaces on the one hand, and enable us to propose more accurate PDE solvers for several fields including engineering, finance and physics on the other hand.
				Prof. Franca Hoffmann	QLA-AIMS			
				Andrès Almansa	University of Paris Cité			
				Julie Delon	University of Paris			

6	Elhadji Moustapha Seck	M	Senegal	Gervais Mendy Bubacarr Bah Jared Tanner	Ecole Superieur Polytechnique / Universite Cheikh Anta Diop AIMS - South Africa and Stellenbosch University University of Oxford	UNIVERSITE CHEIKH ANTA DIOP DE DAKAR	Deep Transfer Learning	Deep learning has recently acquired significant research interest and has been used in many real-world applications. One approach of traditional deep learning methodologies is that the training and test data are drawn from the same domain, so that the input space and data distribution characteristics are the same. However, there are situations where this assumption does not hold. For example, there are cases where one has insufficient training data or difficulties to collect new data. But one of the most powerful idea in deep learning is that sometimes we can take knowledge that the neural network has learned from one task and apply that knowledge to a separate task. For example, one can train a neural network to recognize objects like cats and then use that knowledge or part of that knowledge to help do a better job reading X-ray scans. This is called deep transfer learning. This has motivated many research works that have been done to address deep transfer learning challenges. However, many more important research issues such as how to avoid negative transfer learning remain challenging and need to be understood. This is the main motivation of this PhD project. As most of these researches focused on supervised learning, in this work we are interested in investigating both theoretically and computationally semi-supervised and unsupervised negative transfer learning using deep neural networks. These techniques would allow us to be able to use a mix of labeled and unlabeled data. Simulations will be used to support theoretical findings.
7	Jean Claude TUMUSHIME	M	Rwanda	Prof David Kelvin Dr. Udahemuka Jean Claude Dr. Pacifique Ndishimye	University of Rwanda University of Rwanda AIMS	University of Rwanda	Comparative host immune responses across different hosts and Genomic characterization of Rift Valley fever virus in Rwanda	Rift Valley Fever is an emerging zoonotic mosquito-borne disease caused by Rift Valley Fever virus (RVFV), which infects domesticated ruminants and humans. Rwanda is a country at risk of RVF outbreaks, given its geographical location in the Great Rift Valley with a tropical climate that offers optimal ecological conditions for mosquito proliferation. In addition, some of Rwanda's current development programs such as the rapid expansion of irrigated land initiatives to increase the number of livestock-owning households, conversion of marshlands/wetlands into rice farms and the regional integration have also increased the country's vulnerability to RVF. The first official outbreak occurred in 2018, then 2020 and 2022. The ultimate goal of this research program is to provide real-time data and solutions that can be used by public health and animal health agencies to mitigate RVF outbreaks and epidemics. The study will clearly identify the strains/variants of RVFV that are always behind the recurring RVF outbreaks, shed light on climatic/environmental factors that facilitate the emergence/re-emergence of RVF, determine the socio-economic burden posed by RVF, determine the host immune resilience/vulnerability vis-à-vis RVF infection/disease and propose timely and robust multi-sectoral countermeasures to mitigate RVF outbreaks in Rwanda.
8	Brigitte UMUTONI	F	Rwanda	Prof David Kelvin Dr. Cameline Orlendo Dr. Pacifique Ndishimye	University of Rwanda University of Rwanda AIMS	University of Rwanda	Data-Driven Modelling for evaluating the impact and early detection of Rift Valley Fever outbreak in Rwanda	Rift Valley Fever Virus (RVFV) is a mosquito-borne arbovirus that causes a zoonotic disease in domesticated ruminants and humans. RVFV was first isolated in the Great Rift Valley in Kenya in 1930, where periodic outbreaks continue to pose a significant health threat mainly to countries in East and Sub-Saharan Africa and the Arabian Peninsula. The virus is predominantly harbored and transmitted by mosquitoes, including the Aedes, Culex, Anopheles, and Mansonia genera; however, RVFV may also be spread via contact with the blood, bodily fluids, or tissues of infected animals. RVFV has over 100 known arthropod vectors, with many presents in Europe and the Americas, representing a global threat of the virus emerging in non-endemic regions. Rwanda, as several East African countries, has been experiencing larger and more frequent RVF outbreaks since 2016. The impact of these outbreaks is significant, not only due to the direct consequences of human infections, but also due to the disruption in the use of domestic ruminants as a food source. The main goals of this study are to examine the association of RVFv infection with miscarriage in animals and humans and to establish an early warning system for identifying RVFv outbreaks based on reported abortions.
9	N'guessan Landry Regis Gnankan	M	Cote D'ivoire	Prof Simon King Prof Thomas Niesler Prof Christopher Thron	University of Edinburgh University of Stellenbosch Texas A&M University	University of Stellenbosch	Adapting state-of-the-art neural Text-to-speech models for low-resource languages	n the area of Natural Language Processing (NLP) and Human Computer Interaction (HCI), Text To Speech (TTS) synthesis has received a significant boost in recent years, especially with the development of various deep learning techniques capable of generating excellent quality speech. However, deep learning demands a vast quantity of data, with synthesis systems requiring more than thousand hours of aligned corpus to train end-to-end systems. While a large amount of data is available in open-source for high resource languages such as English, Spanish, and Italian, the majority of languages like Asian and African have little to no training data. But Collecting such an amount of data is a significant obstacle in the case of underrepresented or low-resource languages, for example, the African continent has nearly 2000 languages and dialects and these are mainly spoken and very little written. To develop good speech synthesis systems in these low resource languages, we need innovative ideas. Among these, the development of new neural architectures capable of producing efficient results with a small amount of paired text-speech data would be an appropriate response to the problem of regional languages and dialects that suffer from a significant lack of data and therefore cannot be modeled with classical TTS architectures(Voracious in data) hence the delay in including these languages and the populations that speak them in high-impact technological development tools like voice assistants
10	SODE Akoeugnigan Idelphonse	M	Benin	Prof Dr Romain Glèlè Kakai Prof Peter Diggle Prof Belarmain Adandé Fandohan Dr Elias T. Krainski	University of Abomey-Calavi (UAC), Republic of Benin Lancaster University Université d'Abomey-Calavi (UAC), Bénin King Abdullah University of Science and Technology (KAUST)	University of Abomey-Calavi (UAC), Bénin	Bayesian spatial fusion model framework for joint analysis of point and areal data with biased sampling and correlated outcomes: ecological and epidemiological applications	The availability of spatial data increased dramatically in recent years in many research fields, including epidemiology, ecology, environmental science, remote sensing and economics motivating the use of spatial modelling approaches. The three types of spatial data (geostatistical, areal or point pattern data) usually may come from multiple sources and require the use of spatial modeling framework for their joint analysis to improve parameters estimation and prediction accuracy. Though some studies have explored this research areas in recent years, there still are some methodological scopes regarding potential confounding factors including sampling bias and dependence between the conditional distribution of different outcomes. We propose in this PhD project an extension to spatial fusion modeling to account for these factors. In the first objective, we will make a systematic review with a meta-analysis on the existing spatial fusion models and their application in ecological and health settings. In the second and third parts of our work, models accounting for sampling bias and dependence between conditional distribution of misaligned data will be proposed based on the bias definition and existing multivariate spatial modeling approaches. The performance of our approaches will be evaluated via extensive simulations. The last objective will be constituted of applications in real-world data collected in ecological and epidemiological settings in West Africa.

11	Mamadou DIALLO	M	Senegal	Issa Karambal Stephen E. Moore Philipp Christian Petersen	QLA-AIMS University of Cape Coast/Ghana University of Vienna/Austria)	To be confirmed		
12	Jeremiah Ayock Ishaya	M	Nigeria	Caspar Addyman David Stern Mark Tomlinson	Stellenbosch University IDEMS International Stellenbosch University		Measuring responsive care-giving using machine learning	Responsive care-giving is essential for good health, social, and cognitive outcomes in infancy and early childhood. Consequently, the WHO's Nurturing Care Framework for 2018 positions responsive care-giving as one of five priority areas for promoting early development. High-quality caregiver-infant interactions are essential to social, cognitive, and emotional development. Extensive research has shown that partner synchrony and responsiveness are key indicators of responsive care-giving. Assessing responsive care-giving is key to evaluating the effectiveness of early childhood interventions. But progress in this regard at any level of scale has been slow as it is expensive and time-consuming. Currently, the assessment of early parent-child interactions in low-resource settings is severely hampered by their time-intensive and costly nature. The intent of this study is to develop a new measure of playful parenting that builds on the existing gold standard measures of responsive care-giving combined with state-of-the-art machine learning algorithms and statistical methods. There are some main challenges in building such a tool. First, the data extracted from the interaction using OpenPose is untidy; it contains missing values, noise is introduced, and the process is biased. Secondly, the different statistical approaches used in evaluating and measuring responsiveness are unable to quantify higher-level details such as smiling activity and are also point estimates.
13	Irene Uwerikowe	F	Nigeria	Caspar Addyman David Stern Mark Tomlinson	Stellenbosch University IDEMS International Stellenbosch University		Analysis of parent child interaction videos to identify patterns of different interaction types using machine learning tools.	High quality parent-child interaction is crucial for a child's social, cognitive and emotional development. Children living in low and middle income countries (LMICs) especially those living in sub-Saharan Africa and Asia are at risk of problems in their social, cognitive and behavioral development. Book-sharing (BS) is an interactive parent-child reading that stimulates cognition in infants and young children. Direct observation tools are used to objectively assess behaviors between parent-child interaction to identify strengths and problems that impact on a child's developmental process. Existing methods for assessing these behaviors often make use of manual coding which is time consuming, requires expensive training of skilled professionals to establish reliability, cost effective and labor intensive. In this project, we will develop a novel computational method that is a combination of state-of-the-art machine learning tools and statistical methods to predict interactions between parent-child dyad during shared book reading. This new algorithm will build on existing machine learning tools by modifying them to incorporate statistical methods that can successfully predict these indicators.