ICCBDC

2024 8th International Conference on Cloud and Big Data Computing

ICIIP

2024 13th International Conference on Intelligent Information Processing
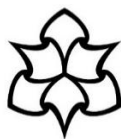
Oxford Brookes University, UK
August 15-17, 2024

Co-Sponsored by

OXFORD
BROOKES
UNIVERSITY

BIRMINGHAM CITY
University

Partners

**Manchester
Metropolitan**
University

UNIVERSITY
of DERBY

GAZİ ÜNİVERSİTESİ
1926

TABLE OF CONTENTS

CONFERENCE COMMITTEES

Conference Chairs

Hong Zhu, Oxford Brookes University, United Kingdom

Huseyin Seker, Birmingham City University, United Kingdom

Program Chairs

Yanlong Zhang, Manchester Metropolitan University, United Kingdom

Mak Sharma, Birmingham City University, United Kingdom

Fatih Kurugollu, University of Sharjah, UAE

Mahir Dursun, Gazi University, Turkey

Publicity Chairs

Alessio Faccia, Coventry University, United Kingdom

Goksel Misirli, Keele University, United Kingdom

Serkan Ayvaz, Bahcesehir University, Turkey

Technical Committee

Mohammad Hasan, Staffordshire University, UK

Amro Al-Said Ahmad, Keele University, UK

Jyh-Haw Yeh, Boise State University, USA

Abhimanyu Mukerji, Amazon, USA

Bedir Tekinerdogan, Wageningen University & Research, The Netherlands

Pascal Lorenz, University of Haute Alsace, France

Yijun Bei, Zhejiang University, China

Ming Li, Wuhan University, China

Man Fung LO, The University of Hong Kong, Hong Kong SAR, China

Roslinazairimah Zakaria, Universiti Malaysia Pahang, Malaysia

Goranka Stanić, School of Art and Design, Croatia

Rajiv Pandey, Amity University Lucknow UP INDIA, India

Bernhard Bauer, University of Augsburg, Germany

Thomas Fraunholz, Smart Cyber Security, Germany

Alberto Trombetta, Insubria University, Italy
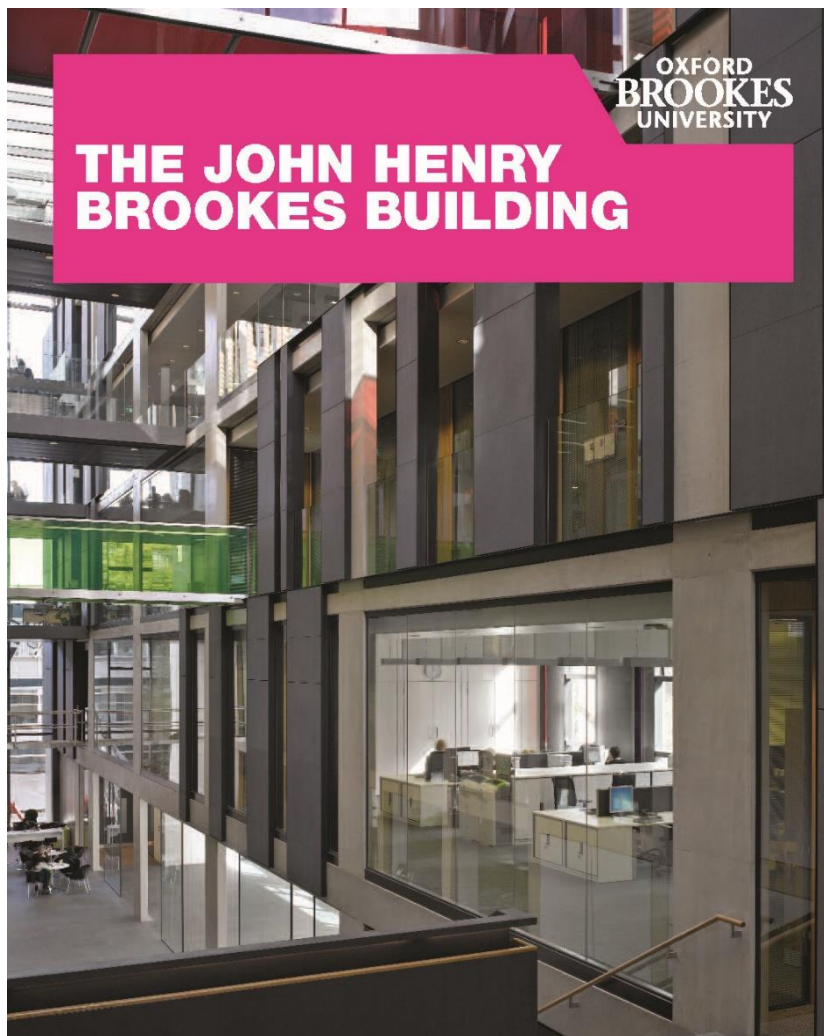
Sima Sinaei, RISE Research Institute of sweden AB, Sweden

Gabriel Gomes de Oliveira, University of Campinas (Unicamp), Brazil

CONFERENCE VENUE

OXFORD BROOKES UNIVERSITY

Address: Oxford Brookes University, Headington Campus, Oxford OX3 0BP, UK

Website: https://www.brookes.ac.uk/

Overview
Headington Campus is home to a thriving student community living and learning in a modern, uplifting environment across three sites.

The buildings on campus house first-rate academic and leisure facilities including collaborative work spaces, the main University library, several catering outlets and a sports centre. The student support team, based in the John Henry Brookes Building, ensures that students have access to a full range of support and advice services on campus.

Direction and Map
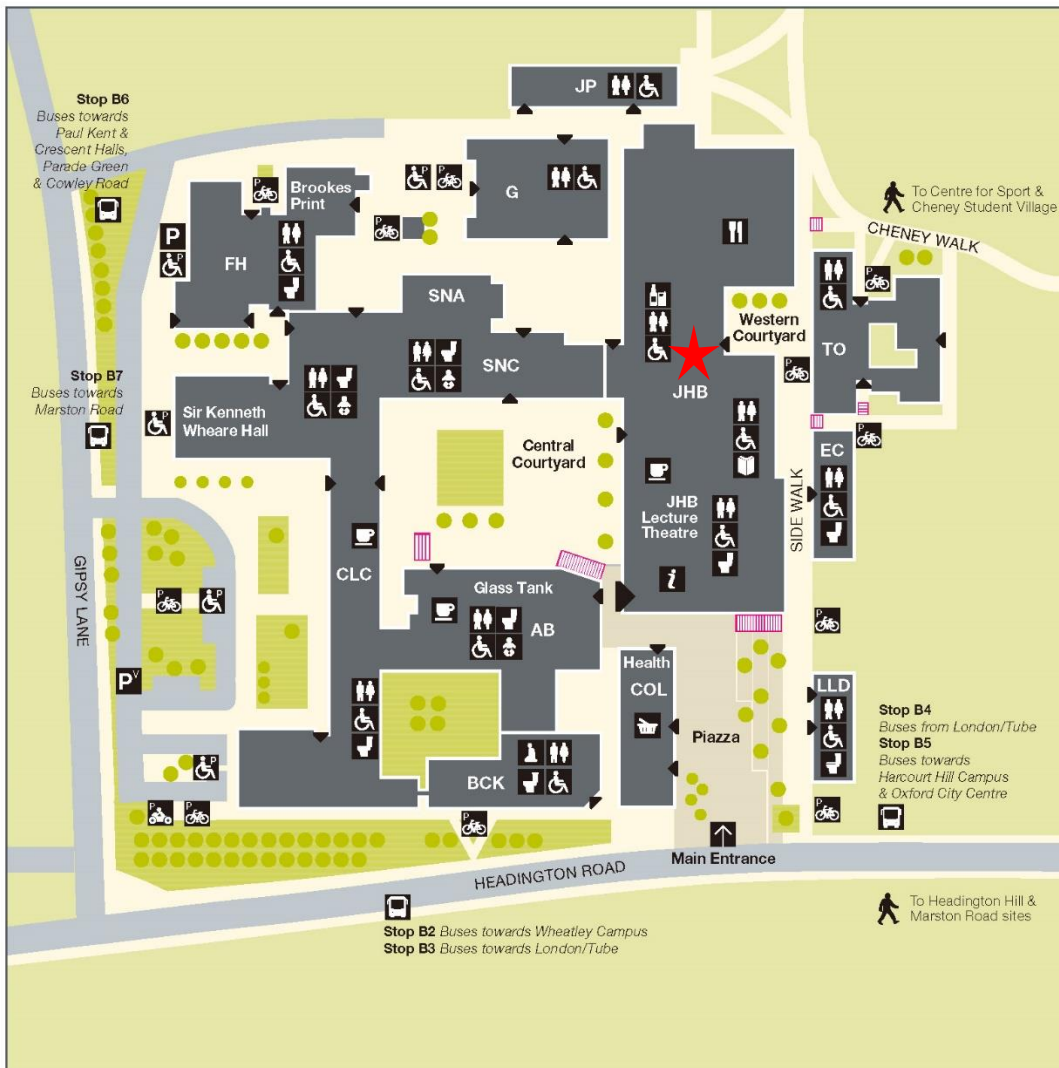
HEADINGTON CAMPUS

GIPSY LANE

AB. Abercrombie
Avo Café
Architecture
Built Environment

BCK. Buckley
Academic Office
Finance & Legal Services
Marketing & Communications
Multifaith Prayer and Quiet Room
Research & Business Development
Strategic Change & Planning
Student Finance
Wellbeing

CLC. Clerici
Business School
Clerici Café
Sir Kenneth Wheare Hall
Vice-Chancellor's Office

COL. Colonnade
Colonnade Store
Health & Life Science
OxINMAHR
Studental

EC. Enterprise Centre
Enterprise Centre
ScaleBase

FH. Fuller
Print, Post & Distribution

G. Gibbs
Humanities & Social Sciences

JHB. John Henry Brookes Building
Main Reception
Brookes Union
Café Central
Careers Centre
Deli
IT Service Desk
Library
Student Central
Taylor's Yard
The Terrace
Centre for Academic Development
24-hour computer rooms

JP. John Payne
Architecture workshops
Estates & Campus Services
Technology Laboratory
Sports & Coaching

LLD. Lloyd
Accommodation Bureau

SNC/SNA. Sinclair & Annex
Health & Life Sciences

TO. Tonge
Bioinnovation Hub
English & Modern Languages
History, Philosophy & Culture

Legend:
- i Information
- Library
- Restaurant
- Cafe
- Bar
- Shop
- Multifaith Prayer and Quiet Room
- Toilets
- Accessible Toilets
- Gender Neutral Toilets
- Baby change
- Bus Stop
- P Parking
- Accessible Parking
- Cycle Parking
- Steps

GENERAL INFORMATION

Oral Presentation Instruction

1. Timing: a maximum of 15 minutes total, including speaking time and discussion. Please make sure your presentation is well timed. Please keep in mind that the program is full and that the speaker after you would like their allocated time available to them.

2. You can use USB flash drive (memory stick), make sure you scanned viruses in your own computer. Each speaker is required to meet her/his session chair in the corresponding session rooms 10 minutes before the session starts and copy the slide file(PPT or PDF) to the computer.

3. It is suggested that you email a copy of your presentation to your personal inbox as a backup. If for some reason the files can't be accessed from your flash drive, you will be able to download them to the computer from your email.

4. Please note that each session room will be equipped with a LCD projector, screen, point device, microphone, and a laptop with general presentation software such as Microsoft PowerPoint and Adobe Reader. Please make sure that your files are compatible and readable with our operation system by using commonly used fronts and symbols. If you plan to use your own computer, please try the connection and make sure it works before your presentation.

5. Videos: If your PowerPoint files contain video clips please make sure that they are well formatted and connected to the main files.

6. Certificates will be given by the session chair right after each presentation.

Dress Code
Please attend the conference in formal attire.

Conference Photos
All the conference photos will be available for download through conference website within one week after the conference.

Safety Reminder: Secure Valuable Items at All Times
We remind you to secure your personal belongings at all times.
* If you are using a laptop computer, do not leave it unattended at any time.
* Keep your purse, wallet and other valuables with you at all times.
The conference organizer will not be responsible for the loss or damage to any personal belongings.

SCHEDULE AT A GLANCE

August 15th (Thursday)

Entrance, John Henry Brookes Building	
10:00-12:00 13:30-17:00	Participants Registration & Conference Kits Collection

August 16th (Friday)

Floor 2, JHB201, John Henry Brookes Building		
Host	Dr. Yanlong Zhang, Manchester Metropolitan University, UK	
9:30-9:35	Opening Remark	Prof. Hong Zhu, Oxford Brookes University, UK
9:35-9:40	Welcome Address	Prof. Huseyin Seker, Birmingham City University, UK
9:40-10:25	Keynote	Prof. Giancarlo Fortino, University of Calabria, Italy *Topic: Integrating Machine Learning and Multi-Agent Systems for Fully Enabling Device-Edge-Cloud Continuum in Complex IoT Worlds*
10:25-10:45	Coffee Break and Group Photo	
10:45-11:30	Keynote	Prof. Hong Zhu, Oxford Brookes University, UK *Topic: User Centric Evaluation of Large Language Models as Program Code Generation Tools*
11:30-12:15	Keynote	Prof. Huseyin Seker, Birmingham City University, UK *Topic: The Power of Data and The Things It Empowers*
12:15-13:30	Lunch	
13:30-15:30	Session 1 - Data Modeling and Machine Learning	
15:30-15:45	Coffee Break	
15:45-17:30	Session 2 - Image Processing and Information Security	
17:45-18:30	Fork Buffet	

August 17th (Saturday)

Activity	
10:30-12:00	Oxford Walking Tour for Voluntary Participants

KEYNOTE SPEAKER

August 16th | 9:40-10:25
JHB201, John Henry Brookes Building

Prof. Giancarlo Fortino
Fellow of IEEE

University of Calabria, Italy

Bio: Giancarlo Fortino (IEEE Fellow 2022) is Full Professor of Computer Engineering at the Dept of Informatics, Modeling, Electronics, and Systems of the University of Calabria (Unical), Italy. He received a PhD in Computer Engineering from Unical in 2000. He is also distinguished professor at Wuhan University of Technology and Huazhong Agricultural University (China), high-end expert at HUST and NIST (China), senior research fellow at the Italian ICAR-CNR Institute, CAS PIFI visiting scientist at SIAT – Shenzhen, and Distinguished Lecturer for IEEE Sensors Council. He was also visiting researcher at ICSI, Berkeley (USA), in 1997 and 1999 and visiting professor at Queensland University of technology in 2009. At Unical, he is the Rector's delegate to Int'l relations, the chair of the PhD School in ICT, the director of the Postgraduate Master course in INTER-IoT, and the director of the SPEME lab as well as co-chair of Joint labs on IoT established between Unical and WUT, SMU and HZAU Chinese universities, respectively. Fortino is currently the scientific responsible of the Digital Health group of the Italian CINI National Laboratory at Unical. He is Highly Cited Researcher 2020-2023 in Computer Science by Clarivate. He had 25+ highly cited papers in WoS, and h-index=81 with 23000+ citations in Google Scholar. His research interests include wearable computing systems, e-Health, Internet of Things, and agent-based computing. He is author of 650+ papers in int'l journals, conferences and books. He is (founding) series editor of IEEE Press Book Series on Human-Machine Systems and EiC of Springer Internet of Things series and AE of premier int'l journals such as IEEE TASE (senior editor), IEEE TAFFC-CS, IEEE THMS, IEEE T-AI, IEEE IoTJ, IEEE SJ, IEEE JBHI, IEEE SMCM, IEEE OJEMB, IEEE OJCS, Information Fusion, EAAI, etc. He chaired many int'l workshops and conferences (130+), was involved in a huge number of int'l conferences/workshops (700+) as IPC member, is/was guest-editor of many special issues (80+). He is cofounder and CEO of SenSysCal S.r.l., a Unical spinoff focused on innovative IoT systems, and recently cofounder and vice-CEO of the spin-off Bigtech S.r.l, focused on big data, AI and IoT technologies. Fortino is currently AVP of the Cybernetics area of the IEEE SMCS and former member of the IEEE SMCS BoG and former chair of the IEEE SMCS Italian Chapter.

Integrating Machine Learning and Multi-Agent Systems for Fully Enabling Device-Edge-Cloud Continuum in Complex IoT Worlds

Abstract: Recently the device-edge-cloud paradigm is gaining momentum due to the benefits it could provide for the development of highly effective, efficient, and complex IoT ecosystems of diversified scale. However, there are many issues related to unsupervised control aspects that need to be addressed in order to fully realize the approach and make it fully operative in real complex environments. In order to address such issues, in this talk, we propose an holistic integration of machine learning and multi-agent systems to create a data-driven control architecture capable to autonomically monitor and control the device-edge-cloud continuum. This objective is being developed in the context of the Horizon Europe project named MLSysOps (https://mlsysops.eu/). Some use cases will be proposed to elucidate our current findings.

KEYNOTE SPEAKER

August 16th | 10:45-11:30
JHB201, John Henry Brookes Building
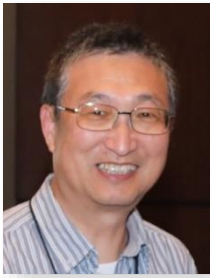
Prof. Hong Zhu

Oxford Brookes University, UK

Bio: Dr. Hong Zhu is a professor of computer science at the Oxford Brookes University, Oxford, UK, where he chairs the Cloud Computing and Cybersecurity Research Group. He obtained his BSc, MSc and PhD degrees in Computer Science from Nanjing University, China, in 1982, 1984 and 1987, respectively. He was a faculty member of Nanjing University from 1987 to 1998. He joined Oxford Brookes University in November 1998. His research interests are in software development methodologies, including software engineering of cloud-native applications, software engineering of AI and machine learning applications, formal methods, software design, software testing, programming languages, software modelling, and automated software engineering tools and environments, etc. He has published 2 books and more than 200 research papers in journals and international conferences. He is a senior member of IEEE, a member of British Computer Society, and ACM.

User Centric Evaluation of Large Language Models as Program Code Generation Tools

Abstract: With the rapid advance of machine learning (ML) technology, large language models (LLMs) are increasingly explored as an intelligent tool to generate program code from natural language specifications. However, existing evaluations of LLMs have focused on their capabilities in comparison with humans. It is desirable to evaluate their usability when deciding on whether to use a LLM in software production. To meet this requirement, we propose a user centric method for this purpose. It includes metadata in the test cases of a benchmark to describe their usages, conducts testing in a multi-attempt process that mimics the uses of LLMs, measures LLM generated solutions on a set of quality attributes that reflect usability, and evaluates the performance based on user experiences in the uses of LLMs as a tool.

We will also report a case study with the method in the evaluation of ChatGPT's usability as a code generation tool for the R programming language. Our experiments demonstrated that ChatGPT is highly useful for generating R program code although it may fail on hard programming tasks. The user experiences are good with overall average number of attempts being 1.61 and the average time of completion being 47.02 seconds. Our experiments also found that the weakest aspect of usability is conciseness, which has a score of 3.80 out of 5.

KEYNOTE SPEAKER

August 16th | 11:30-12:15
JHB201, John Henry Brookes Building

Prof. Huseyin Seker

Birmingham City University, UK

Bio: Huseyin is a research-oriented and enterprise-focused academic and manager with both academic and industry experiences in data science, artificial intelligence, machine learning, and emerging & disruptive technologies/systems. He has published over 100 peer-reviewed articles and has been involved with a portfolio of collaborative research, enterprise and teaching/learning projects of over £ 20M as PI, Co-I and international researcher in collaboration with universities and companies in the UK and abroad. He was one of the founding members of the Institute of Coding and led its activities at Northumbria University until September 2019. He is currently working as Professor of Computing Sciences and Associate Dean (Research and Enterprise) for the Faculty of Computing, Engineering and the Built Environment at Birmingham City University.

The Power of Data and The Things It Empowers

Abstract: Data has become an asset in every sector and discipline. With technological advancements and widespread access to the internet and commonly used connected devices in our daily lives (e.g., mobile phones, social media, the internet of things), we have become data producers, contributing to the collection of data sets. The collection and analysis of data using advanced data analytics technics and artificial intelligence methods have started making a positive impact in society, economy and environment around the world. However, we have also started seeing unethical use of such data sets. Due to some recent disturbing examples, policymakers need to regulate the data market to make our digital world safer for current and future generations. This talk will cover both aspects of the use of data-driven artificial intelligence methods along with examples and developments with future direction.

Session 1 - Data Modeling and Machine Learning
Chairperson: Prof. Hong Zhu, Oxford Brookes University, UK

Date: August 16th, 2024	Room: Floor 2, JHB201

*Note: The schedule of each presentation is for reference only. Authors are required to attend the whole session, in case there may be some changes on conference day. Please join in the room 5-10 minutes earlier.

DC24-322

13:30-13:45

Dementia Deterioration Prediction Using Machine Learning

Layla Dawood Almardoud, **Hissam Mouayad Tawfik**, Sohaib Majzoub
University of Sharjah, United Arab Emirates

Abstract: Dementia is a disease that imposes medical, social, and economic challenges on medical professionals, caregivers, and the patients themselves. Dementia monitoring and prognosis are critical factors besides dementia diagnosis. However, recent studies on dementia prognosis involve people with diagnosed dementia and not non-demented with a high risk of being demented, like people with cognitive difficulties. The aim of this paper is to use Machine Learning (ML) algorithms to predict patients with the risk of deterioration from medical histories containing clinical, cognitive, and profile data collected from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The best model was the Random Forest model with a sensitivity of 0.79, accuracy of 0.77, specificity of 0.76, F1-score of 0.78, and an AUROC of 0.83. Moreover, the model was interpreted through permutation importance. Using the permutation importance tool, the study highlighted the strong effect of diagnosis information and specific symptoms like muscle pain for dementia deterioration prediction.

DC24-328

13:45-14:00

Integrating Federated Learning and Differential Privacy for Secure Anomaly Detectionin Smart Grids

Mohammadreza Mohammadi, Rakesh Shrestha, **Sima Sinaei**
RISE Research Institute of Sweden, Sweden

Abstract: Anomaly detection is essential for ensuring the safe and efficient operation of industrial systems like smart grids. Smart grid stations handle sensitive data and are often hesitant to share it with third-party servers for centralized anomaly detection. Federated Learning (FL) offers a viable solution to this issue by enhancing anomaly detection in smart grids without compromising data privacy. We present a method for developing an unsupervised anomaly detection system using FL applied to a synthetic dataset that mimics a real-world smart grid system's behavior. We focus on utilizing FL's long short-term memory autoencoder in short, LSTM-AE for anomaly detection. However, there are concerns about potential privacy breaches in the FL system. Hence, to address this issue, we propose to integrate differential privacy (DP) with FL for anomaly detection by adding artificial noise to parameters at the client side before aggregation. This method ensures data privacy while maintaining the convergence of federated learning algorithms. Moreover, this research determines the optimal privacy level to balance noise scale and model accuracy. Our findings suggest a criterion for selecting the right privacy budget of DP based on the requirement of the system to provide good level of privacy in the system while maintaining the f1-score of FL-based anomaly detection system greater than 90%.

DC24-411A

14:00-14:15

Research on Automatic Guidelines Reference Extraction and Mining Based on Representation Learning

Huijie Han, Shiqi Guo, Xiaocui Gong and Xinying An
Chinese Academy of Medical Sciences & Peking Union Medical College, China

Abstract: The identification of cited references in clinical guidelines is crucial for tracking the application of academic research to clinical practice. To tackle the problem of excessive references in clinical guidelines, this study aims to accurately identify and extract the characteristic elements of cited references to facilitate in-depth analysis and exploration of clinical guideline citations. In this paper, we propose a new method to

extract the references from clinical guidelines.Leveraging full-text papers from PubMed as the data source, we utilize the Doc2Vec representation learning method to obtain semantic vector representations of clinical guidelines' references. Data annotation is carried out semi-automatically using the Albert deep learning model. Furthermore, we utilize semantic vector representation of metadata, such as guidelines references titles, achieved through representation learning methods. Experimental results demonstrate that the trained automatic identification model outperforms other deep learning models in effectively recognizing references cited in clinical guidelines.The integration of advanced representation learning techniques with deep learning models in this study showcases a sophisticated and robust framework for the precise extraction and semantic analysis of cited references in clinical guidelines, paving the way for enhanced understanding and utilization of academic research for clinical practice.

DC24-310

14:15-14:30

RECAST: An Open Source Platform for Item-Specific Capturing of Real Production Processes

Tim Köhler, **Thomas Fraunholz**, Dennis Rall, Lars Larsen, Dominik Görick, Alfons Schuster
German Aerospace Center (DLR), Germany

Abstract: The digitization of production is an essential component of Industry 4.0. The collection of data in the production process is a decisive factor and a prerequisite for the use of artificial intelligence. At machine level, groundbreaking successes have been achieved in the field of predictive maintenance using machine learning, which is directly reflected in increased productivity and cost savings. In contrast, digitization at item level, i.e. the digital tracking of all items in the process context that ultimately lead to the finished product, is much more complex. A non-structured collection of data at item level prevents the use of artificial intelligence, for example in the quality testing of products. As part of the RECAST project, the extent to which existing non-networked production facilities can be digitized with this objective was investigated. As part of the work, a framework was created using exclusively open source components, which allows a digital representation of various real existing complex production processes in the aerospace industry to be created. With this digital twin, the system data can be recorded in real time on a item-specific basis. Thanks to the generic approach to process mapping, the results can be applied to a wide range of application fields in mechanical engineering and data-intensive administrative tasks.

DC24-313

14:30-14:45

Purity: a New Dimension for Measuring Data Centralization Quality

Lander Bonilla, María José López Osa, Josu Diaz-de-Arcaya, Ana I. Torre-Bastida, Aitor Almeida
Tecnalia, Basque Research & Technology Alliance (BRTA), España

Abstract: Data has become an asset for companies, originating from various sources, such as IoT paradigms. It is crucial to safeguard its life cycle using suitable, scalable, and effective technologies, like those enabled by cloud computing models. However, in order to extract value from this data, complementary processes of collection, refinement, cleaning, or modeling, among many others, are required. Furthermore, organizations greatly vary in their methodologies and approaches to handling data, which further emphasizes the need for standardized techniques. In this regard, data management methodologies promote the adoption of the various dimensions of data quality in order to ensure the reliability of data across different systems and processes. The main contribution of this manuscript is the proposal of a new data quality dimension, coined purity, to measure the importance of the data in a processing pipeline topology. As a result, organizations can better guarantee the quality of their datasets in order to raise the success of data-driven endeavors within organizations. The proposed methodology is validated in an urban mobility use case.

DC24-316

14:45-15:00

Interpretable RNN for Prediction & Understanding of Childhood Obesity: A Scenario from the UK Millennium Cohort Study

Hissam Tawfik, **Balbir Singh**, Tarek Khater
Leeds Beckett University, United Kingdom

Abstract: Childhood obesity is a growing public health concern worldwide, with significant implications for long-term health outcomes and healthcare costs. Early identification of children at risk of obesity is essential for implementing effective prevention and intervention strategies. In recent years, machine learning techniques have emerged as powerful tools for predicting childhood obesity risk based on various predictors such as demographic factors, dietary habits, physical activity levels, and genetic predisposition. This paper explores the use of temporal deep-learning models, specifically recurrent neural networks and long short-term memory, to predict obesity childhood stages using longitudinal datasets. The long short-term memory model emerges as the top performer, achieving an accuracy of 85% and an F1-score of 0.85, surpassing existing models. To enhance model transparency, model-agnostic methods are employed, revealing critical insights into model predictions. Results indicate that BMI at age 11 is the most influential feature in predicting obesity. These findings underscore the importance of applicability models and explainable machine learning methods in understanding and predicting childhood obesity progression.

DC24-302

15:00-15:15

Optimality Analysis of a Two-Server Queuing System with the RID Discipline

Xian Liu, **Changcheng Huang**
Carleton University, Canada

Abstract: Latency, jitter, and reliability are three KPIs in QoS of modern edge and cloud computing systems. Their analytical insights can be well gained through queuing theory. In this paper, we develop a queuing model with two parallel heterogeneous servers with a randomly-initial-dispatch discipline. This model generalizes several existing models, yet simple enough as it can be configured by a couple of parameters. With this simplicity, we analytically prove the optimality of several standard measures of effectiveness for the latency and jitter.

DC24-323

15:15-15:30

Analyzing the Usability, Performance, and Cost-Efficiency of Deploying ML Models on BigQuery ML and Vertex AI in Google Cloud

Hongyu Wang, **Jeong Yang**, Gongbo Liang, Young Lee, Zechun Cao
Texas A&M University-San Antonio, United States

Abstract: This study compared and analyzed the usability, performance, and cost-efficiency of deploying Machine Learning (ML) models in two ML-AI platforms in Google Cloud: BigQuery ML and Vertex AI. Through the experiments with two separate cases, the analysis was conducted with MIMIC-IV datasets of hospitalized patients to deploy regression models on each platform to predict mortality and progression of diseases. The documentation, learning curve, and resource suitability of the platforms were evaluated to access their usability. The study evaluated the total running times and resource utilizations, including storage and compute, to analyze their performance and cost efficiency. The analysis results showed that BigQuery ML offers good usability with easy-to-follow documentation and a moderate learning curve for cloud users, making it more suitable for SQL-savvy users and large-scale data analytics tasks. It also showed efficient resource management and deployment despite its higher initial processing times during the training. Vertex AI incurred higher costs due to longer training times and specific resource allocations. The findings indicate that BigQuery ML seems to be more efficient, particularly in terms of processing time and cost for the experimented clinical dataset and regression models, emphasizing its suitability for large-scale data processing tasks where efficiency is essential.

Best Presentation Award & Session Group Photo

Session 2 - Image Processing and Information Security
Chairperson: Prof. Huseyin Seker, Birmingham City University, UK

Date: August 16th, 2024	Room: Floor 2, JHB201

*Note: The schedule of each presentation is for reference only. Authors are required to attend the whole session, in case there may be some changes on conference day. Please join in the room 5-10 minutes earlier.

DC24-420

15:45-16:00

An Investigation into the Application of Anomaly Detection and the Meijering Filter in the eKYC Process to Detect Recaptured Identity Documents

John Magee, Stephen Sheridan, Christina Thorpe
Technological University Dublin, Ireland

Abstract: As financial institutions move to offer more digital based services, the need for secure and accurate eKYC services increases. Identity documents submitted during the eKYC process are fundamental to establish the identity of customers. The ability of bad actors to modify identity documents using digital imaging software opens these eKYC services to new threats, resulting in identity theft and reputational damage. In this exploratory research we apply auto-encoder/decoder anomaly detection to the domain of recaptured identity document detection, using the Meijering filter as the feature extraction mechanism.

DC24-421

16:00-16:15

Automatic Segmentation of Table Tennis Match Video Clips Based on Player Actions for Enhanced Data Acquisition

Zhong-Kai Wei, Jieh-Ren Chang
National Ilan University, Taiwan

Abstract: The collection of data in table tennis competitions is crucial for professional coaches and top players. Big data provides them with insights to analyze opponents' strategies and adjust their own training directions. However, existing models for recognizing table tennis serving and receiving actions still suffer from low accuracy, making them unsuitable for automated match data recording. Currently, training data for machine learning models are sourced from manually edited videos with labels, which is time-consuming, requires high levels of expertise, and generates limited data, leading to poor accuracy in existing action recognition models. Therefore, this study aims to realize automatic video clipping of table tennis matches using computer vision and machine learning techniques to enhance data collection efficiency. Additionally, advanced models for recognizing table tennis serving and receiving actions are developed. These models utilize reinforcement learning and semi-supervised learning techniques, leveraging automatically clipped, unannotated action segments, along with a small amount of annotated data, to enhance accuracy. The proposed Table Tennis Playing Status Recognition Model (TTPSRM) achieved a serve recognition accuracy of 98.4% and a receive recognition accuracy of 96.6%. The automated video segmentation system, tested with 2022 ITTF competition videos, successfully clipped 70% of serve actions, with average start and end point errors of 0.547 and 0.287 seconds, respectively.

DC24-404A

16:15-16:30

Obesity and Pre-Diabetic Influence on the sEMG Signal Classification

Mohamed Rashed Al-Mulla
Kuwait University, Kuwait

Abstract: This research is investigating the influences on the sEMG signal classification from isometric contraction of the biceps brachii in obese and non-obese subjects. Previous research has established a lower classification accuracy among obese subjects, compared to non-obese subjects. This study aims to gain a clearer view of the effects on the sEMG signal in obese participants. In our preliminary trials, surface Electromyography (sEMG) signals were recorded from 14 male subjects performing isometric contraction of the biceps brachii, reaching complete fatigue occurrence. Two groups participated in the study; one group of 7 males were obese, while in the other

group, 7 participants were classified as non-obese. A range of parameters were used to classify the signal, both in the time domain and frequency domain. A range of feature extraction methods were utilised to determine if the classification of non-obese decreased consistently for the various parameters. The preliminary findings show that the classification accuracy decreased for obese subjects compared to non-obese subjects, in a range of feature extraction methods. This suggests there is an element affecting the classification performance of the signal. Although previous research has identified crosstalk as one of the reasons for the lower classification accuracy, other studies are arguing that for diabetic patients, diabetes mellitus (DM) and glycemic control are important factors in muscle fatigue. Additionally, evidence has shown that diabetes risk factors are related to fatigue symptoms. With these findings in mind, this study is investigating whether pre-diabetes is a factor in muscle fatigue detection during isometric contractions, and if so, in what ways will this affect the signal classification. Further studies will be conducted using invasive or non-invasive (although the latter is preferred) methods to determine the glucose level in the studied muscle. A valuation will be carried out to establish if the fat layer itself is affecting the signal, if the glucose level influences the signals or if other factors should be taken into account.

DC24-405A

16:30-16:45

Non-Local Energy Segmentation Model Based on Wasserstein Distance

Qiang Hou, Shangcan Liu, Mingling Ou
University of Geosciences, China

Abstract: The target edges in Medical images affected by imaging devices and organ tissue properties have blurred characteristics, making traditional methods to image segmentation very difficult. To address this challenge, a new model applicable to medical image segmentation with edge blurring aims at improving segmentation accuracy, protecting the geometric structure of edges and improving the robustness of the algorithm. The new model applies the area information to the edge-based active contour model. First, a new Signed Pressure Force (SPF) function was constructed with the global and local information. The local SPF function and the weight function between the length term and the area driver term were constructed using the non-local method and Wasserstein distance. Then the energy generalization function is defined. After that energy function is minimized and the partial differential equation for level set function evolution is established using variational method. Finally, the evolution equation of the level set is solved numerically to segment the interested region of medical images. The results show that the new model can locate the edges better and segment the target more integrally.

DC24-306

16:45-17:00

Assessing Pretrained Model Through Transfer Multi-Task Learn For Melanoma Classification

Hongyuan Xie, Yanlong Zhang
Manchester Metropolitan University, United Kingdom

Abstract: Melanoma is a lethal skin cancer that is increasingly threatening the public health system due to increased incidence rates and mortality rates. Early detection of the disease is vital for improved outcomes and the reduction of mortality rates. Skin cancer classification remains a challenging task in the field of dermatology. While self-attention mechanisms and large language models have gained traction in skin cancer detection research, there is still insufficient evidence demonstrating their superior performance compared to CNNs. Thus, further exploration of this area is warranted. Where the quest for the optimal CNN pretrained model persists. In this study, we address this gap by assessing various pretrained models to determine the most effective one for skin cancer classification. Additionally, we introduce a novel approach that leverages transfer learning to develop a multi-task model capable of providing more comprehensive prediction information from dermatological images. Unlike conventional single output classification tasks that rely solely on label prediction, our proposed model utilizes transfer learning techniques to extract valuable features from pretrained models, enhancing its ability to predict multiple tasks simultaneously. This novel approach not only advances the field of dermatology by improving classification accuracy but also meets the growing demand for more informative predictions in clinical settings.

DC24-314

17:00-17:15

Securing Software Defined Networking in Cloud Infrastructure through Transport Layer Security

Majid Jalali, Martin J Reed
University of Essex, United Kingdom

Abstract: Cloud and big data applications increasingly rely on the softwarization of underlying infrastructure to achieve highly flexible reconfiguration. Software Defined Networking (SDN) stands as a vital component, offering centralized control and efficient orchestration of network resources. This paper underscores the critical importance of security within cloud and big data environments and highlights SDNs role in this context, particularly through the implementation of Transport Layer Security (TLS) encryption. TLS establishes secure communication channels between SDN controllers and switches. While prior studies have noted potential latency impacts of TLS, this paper presents contrasting findings within the context of SDN switches widely used in cloud infrastructure. Through a comprehensive examination, this study shows the pivotal role TLS plays in fortifying the integrity and security of SDN communications. Technical insights into TLS integration, its impact on performance, and practical case studies offer valuable guidance for enhancing the security of SDN networks which play a key role in cloud infrastructures and big data applications.

DC24-324

17:15-17:30

Prediction of Net Effective Wind Pressure in Walls using Artificial Neural Network and Akaike Information Criterion

Dante L. Silva, Kevin Lawrence M. De Jesus, Benjamin D. Flores, Russell L. Diona, Ralph Alwin M. De Jesus, Orlando P. Lopez
Mapua University, Philippines

Abstract: Wind forces on structures have the potential to cause significant damage. A database involving the distance from the ridge, enclosure classification, surface type, elevation above ground level, wind direction, basic wind speed, presence of wall/surface openings, and effective net wind pressure (ENWP) was created using computation fluid dynamics (CFD). This paper focuses on the development of a model for predicting ENWP using a backpropagation-artificial neural network (BP-ANN). Utilizing the Levenberg-Marquardt algorithm (LMA) and hyperbolic tangent sigmoid function (HTSF) as the model hyperparameters, the study investigated several network structures and the simulations revealed that the 7-20-1 is the best model among the topologies observed in this study. The results showed an R value of 0.99868, MSE and MAPE of 0000749 and 5.036%, respectively. Additionally, the Akaike Information Criterion (AIC) was used as another layer of metric to measure the effectiveness of the model. The least was observed in the 7-20-1 network structure indicating that this is the best among the topologies observed in this study. Moreover, a sensitivity analysis (SA) through Garson's Algorithm (GA) was performed to determine the relative contribution (RC) of the input parameters (IP) including the distance from the ridge, enclosure classification, surface type, elevation above ground level, wind direction, basic wind speed, and presence of wall/surface opening to the effective net wind pressure. The findings presented that the basic wind speed is the most significant parameter to the effective net wind pressure value. The results of this study can be utilized in considering appropriate configuration to minimize the effects of wind pressure in structures.

Best Presentation Award & Session Group Photo

Posters

DC24-304

Optimizing Hadoop Distributed File System Replication Policies with Predictive Categorization

Nada A. Zayed, Yasmine N. M. Saleh, Ahmed A. Aboelfarag, Mohamed A. Shaheen
Arab Academy for Science, Technology and Maritime Transport, Egypt

Abstract: The rapid increase in data volume over the past years has given rise to big data science. Distributed File Systems (DFS) have become widely employed to handle this vast amount of data, such as Google File System and Hadoop Distributed File System (HDFS). The primary objective of a DFS is to ensure data availability and system reliability in the event of failure. Data availability and system reliability are achieved by replicating files across multiple locations, which, however, results in the consumption of storage space and other resources. The significance of these files varies based on their frequency of use within the system. As a result, specific files are deemed less critical and do not warrant extensive replication, as they hold little importance in the overall system. This paper presents a novel approach called "Optimizing HDFS Replication Policies with Predictive Categorization" for storage efficiency. This approach aims to minimize storage consumption while ensuring data availability and system reliability. Experimental results using the Spotify Hit Predictor Dataset (1960-2019) showcase significant improvements in storage utilization and throughput. This approach not only enhances system performance but also positions organizations for substantial cost efficiencies in managing their data-intensive workloads.

DC24-318

Blockchain-based Secure Storage and Cross-domain Sharing Mechanism for Medical Image Data

Chuanjia Yao, Rong Jiang, Leijin Long, Jianhang Dong, Chenguang Wang
Yunnan University of Finance and Economics, China

Abstract: With the rapid advancements in medical imaging technologies such as CT, MRI, PET, and ultrasound, these modalities have become pivotal for precise clinical diagnoses and treatment planning. Notably, they facilitate accurate patient evaluations and expedite early disease detection and targeted interventions. However, the voluminous nature of medical image data, coupled with stringent privacy requirements, poses significant challenges to efficient and secure storage and sharing within the realm of big data management. In light of these complexities, this paper introduces a blockchain-centric framework that innovatively integrates InterPlanetary File System (IPFS), cloud storage, AES and blockchain technologies. The framework is meticulously designed to address the dual objectives of secure storage and streamlined sharing of medical image data. A key novelty lies in the strategic differentiation of cloud service roles: the cloud's "chain-on" nodes participate in blockchain's validation and consensus processes without engaging in direct data management or operations, whereas "chain-off" nodes are tasked with establishing   transient channels for data transmission upon successful validation of user requests by their chain-on counterparts. In practice, medical image owners encrypt the data using IPFS and create indexes derived from examination reports, facilitating keyword-based searches for data users. Moreover, a Hybrid Encryption scheme with Multiple Public Keys is implemented, necessitating users to utilize their blockchain management system-generated private keys to unlock the AES decryption key for image data access. This multi-faceted technology integration approach not only fortifies the security of storage and ensures privacy preservation but also streamlines retrieval processes and bolsters permission management for medical image data. Therefore, it presents a pioneering solution for medical image data storage and sharing.

DC24-320

Leveraging Workload Prediction for Query Optimization in Multi-Tenant Parallel DBMSs

Mira El Danaoui, Shaoyi Yin, Abdelkader Hameurlain, Franck Morvan
Paul Sabatier University, France

Abstract: In the realm of Database-as-a-Service (DBaaS), multi-tenancy has emerged

as a key concept, enabling a single database instance to be shared among multiple tenants. Within a multi-tenant parallel DBMS, at any point, many tenants may submit SQL queries concurrently, each associated with a performance Service-Level Objective (SLO). Any violation to these SLOs requires the provider to pay fines. Given the instability of query workload in these systems, along with the need to satisfy the strict performance objectives, innovative query optimization strategies are required. These strategies should be well adapted to the specific needs of multi-tenant environments while ensuring profitability for the service provider. In this paper, we design a workload prediction management system (or Prediction Manager in short) inside the query optimizer. The main objective of this manager is to continuously predict the upcoming workload and make this information available to the optimizer. This enables the query optimizer, along with the performance SLOs, to decide when and which execution plan to select. Experimental results showed high efficiency in terms of the provider's long-term profit relatively to those in existing literature.

DC24-329   Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis

Wenjun Gu, Yihao Zhong, Shizun Li, Changsong Wei, Liting Dong, Zhuoyue Wang, **Chao Yan**
Northeastern University, USA

Abstract: The stock market's ascent typically mirrors the flourishing state of the economy, whereas its decline is often an indicator of an economic downturn. Therefore, for a long time, significant correlation elements for predicting trends in financial stock markets have been widely discussed, and people are becoming increasingly interested in the task of financial text mining. The inherent instability of stock prices makes them acutely responsive to fluctuations within the financial markets. In this article, we use deep learning networks, based on the history of stock prices and articles of financial, business, technical news that introduce market information to predict stock prices. We illustrate the enhancement of predictive precision by integrating weighted news categories into the forecasting model. We developed a pre-trained NLP model known as FinBERT, designed to discern the sentiments within financial texts. Subsequently, we advanced this model by incorporating the sophisticated Long Short Term Memory (LSTM) architecture, thus constructing the innovative FinBERT-LSTM model. This model utilizes news categories related to the stock market structure hierarchy, namely market, industry, and stock related news categories, combined with the stock market's stock price situation in the previous week for prediction. We selected NASDAQ-100 index stock data and trained the model on Benzinga news articles, and utilized Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Accuracy as the key metrics for the assessment and comparative analysis of the model's performance. The results indicate that FinBERT-LSTM performs the best, followed by LSTM, and DNN model ranks third in terms of effectiveness.

DC24-414   IOT-Based Yield Prediction System for Better Vegetable Production Secure and Energy-Efficient Edge Computing Platform with Customized RISC-V

Cuong Pham-Quoc, Nguyen The Binh
Ho Chi Minh City University of Technology (HCMUT), Vietnam

Abstract: Vietnam is experiencing rapid economic growth, but it also has serious environmental problems, most notably increased trash generation and plastic pollution. Biodegradable Garbage Bags have become a viable way to address this problem. But obstacles stand in the way of their widespread acceptance, especially in rural areas. This study explores the variables influencing Vietnamese consumers' opinions and intentions to buy biodegradable garbage bags. The research intends to direct strategies for encouraging sustainable consumption patterns and fostering environmental consciousness

Oxford Walking Tour

There's no better way of exploring over 1,000 years of history in the City of Dreaming Spires than on foot. You will encounter Oxford's past as a Saxon stronghold through to the modern day life of the thriving University city, and all of the fascinating history and stories that lie in between.

Tours will open your eyes to the most famous spots in Oxford, from its historic buildings to iconic colleges, many of which are now represented in films and TV ranging from Harry Potter to Inspector Morse.