AI Satellite Abstracts

AI01: Māori AI & Data Sovereignty: Embedding Values.

Taiuru, K.N.[1]

[1]Taiuru & Associates Limited., NZ.

The presentation addresses the critical issue of ethical and cultural alignment with Māori AI and Data. It highlights the importance of ensuring that AI and data practices are aligned with Māori values and principles to prevent bias and discrimination. The problem is underscored by the current lack of Māori representation in the tech industry, the biases in AI algorithms, and the insufficient inclusion of Māori voices in AI design and development processes.

The findings of the presentation emphasise the need for integrating Te Tiriti values into AI and data governance to ensure that Māori data is treated as a taonga and subject to Māori data governance. The significance of these findings lies in the potential for achieving equitable outcomes for Māori, contributing to Māori development, and preventing digital colonisation.

By recognising Māori data as a taonga and embedding Māori leadership and decision-making at all levels of the AI system's lifecycle, the presentation advocates for a more inclusive and equitable approach to AI and data practices.

AI02: The AI In Health Research Network

Wong, LE.[1], Baker, N.[2], Dobson, R.[2,3]

[1]Strategic Engagement, University of Auckland, Auckland, NZ, [2] School of Population Health, University of Auckland, Auckland, NZ, [3] Planning, Funding and Outcomes, Health New Zealand | Te Whatu Ora

The 2023 Prime Minister's Chief Science Advisor's (PMCSA) AI in Health report[1] highlighted the urgent need for AI digital tools in the New Zealand health sector, emphasising a unified approach to streamline research efforts, foster collaboration, and expedite the translation of research findings into practical solutions for health sector practitioners. A recommendation (R20) from the report is the establishment of a national network for AI research with a specific focus on health. In response to this recommendation, the University of Auckland and MedTechIQ, have provided initial seed funding to facilitate the establishment of an AI in Health Research Network (AIHRN).

The Network seeks to create an environment that enables experts to come together to share knowledge, collaborate, and develop real-world applications to position New Zealand at the forefront of AI health innovation. Through a culture of collaboration and co-development, we aim to facilitate the adoption of AI tools in health to develop safe solutions and innovate much-needed efficiencies to support a system that is under pressure. The network would facilitate better cohesion across academia, as well as industry and health services. The development of these strategic relationships will foster collaboration, sharing of outputs and lessons learned, ultimately strengthening transparency and domestic social license.

AI's potential to revolutionise health outcomes and healthcare delivery in Aotearoa New Zealand and worldwide is well documented. Taking a coordinated, multidisciplinary approach with a cross-section of experts from around the country, and internationally, will allow us to amplify this potential and create an ecosystem with solutions that are adapted to our national circumstances, future trajectory and ultimately protect and preserve the health of New Zealanders.

1. *Capturing the benefits of AI in healthcare for Aotearoa New Zealand.* Office of the Prime Minister's Chief Science Advisor Kaitohutohu Mātanga Pūtaiao Matua ki te Pirimia. October 2023.

AI03: Exploring AI-mediated patient and provider relationships: a study protocol for codesign with Pacific heart failure patients

Hanchard, S.[1], Chung, C.[2], Selak, V.[1], Grey, C.[1]

[1]Faculty of Medical and Health Sciences, The University of Auckland, NZ [2]Department of Accounting and Information Systems, University of Canterbury

Heart failure (HF) is a chronic, relapsing syndrome that responds well to ongoing management in the community, yet there are known gaps in continuity of care following discharge from hospital. Empowering patients to accurately and confidently interpret symptoms and to feel confident when to seek further care can reduce unplanned hospital readmissions. Pacific peoples experience significant disparities in HF outcomes and face additional language and cultural barriers in equitably accessing services to manage their condition. While digital health solutions have already been developed, with and without artificial intelligence (AI) capabilities, to support HF patients in the community, little research has been undertaken on the requirements of Pacific patients and families in using digital tools; and whether digital tools can safely optimise patient and provider engagement and management. The study outlined in this protocol aims to understand the acceptability and feasibility of AI-enhanced support from the perspectives of Pacific HF patients, providers and other system stakeholders and to codesign a culturally-grounded digital health solution.

This qualitative project consists of interviews and workshops and is the first phase within a multi-phase programme. Pacific frameworks, such as Fa'afaletui, and codesign methodologies underpin an approach that upholds the needs and viewpoints of Pacific families in the development lifecycle. Relational worldviews and ethics inform an exploration of how HF patients self-manage their condition in partnership with their providers, and how communication mediated through technology might enhance or detract from those relationships. The overall aim of this research is to strengthen the self-management and engagement of Pacific patients and healthcare providers during post-discharge management, towards improved long-term outcomes for living with HF. Findings from this research may inform the use of AI tools for chronic care management more broadly, particularly for Māori and other underserved populations.

AI04: Artificial Intelligence Research, Key Applications, and Infrastructures at VUW

Zhang, M[1], MacDonall, S.[1], Xue, B.[1], Liu, I.[1], Arnold, R.[1]

[1]Centre for Data Science and Artificial Intelligence, Victoria University of Wellington, Wellington, NZ.

Te Whiri Kawe -- The Centre for Data Science and Artificial Intelligence (CDSAI) at Te Herenga Waka -- Victoria University of Wellington is a university-wide research centre launched by the Minister of Research Science and Innovation of the New Zealand Central Government in June 2023. This talk will provide an overview of research and developments at the Centre with a focus on the capabilities, fundamental research directions and key applications of artificial intelligence and data science. The fundamental research directions will cover modelling and statistical learning, evolutionary learning and optimisation and multi-objective decision making, image, text, signal and language processing, deep learning and transfer learning, interpretable AI and machine learning, generative AI and large language models, multi-modal learning, time series analysis/data stream mining and lifelong learning, scheduling and combinatorial optimisation. They key AI applications will include primary industry, climate change, (bio)medical/health applications, high-tech/high-value manufacturing, and AI ethics and public policies that generate impact to NZ economy, environment, health outcome, and social aspects. We will discuss potential collaboration opportunities with stakeholders in various problem domains. Finally, the infrastructure for AI research and applications will be discussed.

AI05: AI Infrastructure for science: for one, for all

Jones N. D.[1]

[1]Digital Research Innovation & Artificial Intelligence, University of Auckland, Auckland, New Zealand.

The noise about the incessant rise of Artificial Intelligence (AI) is near deafening, whether on the risks to jobs, potential for violations of ethics, or the transformation of industries and emergence of general intelligence. Within this vast space of risks and opportunities sit the aspirations and capabilities of science. While AI has been the provenance of computer scientists for many decades, in recent decades its impact and hence appeal has broadened. A more fundamental shift is occurring in the last two years that is catching even the most advanced research teams by surprise. This raises the promise of AI becoming relevant for a vast array of the sciences. It raises the interests of scientists broadly in accessing infrastructure and capability to make sense of what it might mean in their field. It raises the questions of what infrastructure and capability is available to undertaken scientific research that depends on AI. What do we have available within or via our institutions, and what is accessible nationally and internationally? How are we currently making use of these infrastructures, and where might we be heading? We will look at this landscape of AI infrastructure and capability, and explore some of the insights gained from supporting AI for science nationally and institutionally.

AI06: Whither artificial intelligence: implications for science

Sir Peter Gluckman ONZ KNZM FRS[1]

[1]Koi Tū: Centre for Informed Futures, Auckland, NZ

AI might change the very nature of universities, education and science as well as creating fundamental changes in many sectors including government. But our national investment has been poor and not strategically coordinated. As AI becomes a pervasive technology clearly its implications for the production and reporting of knowledge are enormous. But its implications for how scientific knowledge is received, trusted and used in decision making be it by individuals or governments cannot be underestimated. New Zealanders have relatively low trust in AI. The limitations of popular LLMs are poorly understood. The lack of checks and balances on how AI might be used in matters affecting citizens are factors in how trust in institutions including science might evolve. AI already has demonstrated enormous uses in science allowing investigations of data in ways never before possible. It will change the way much science is done. But poor research using AI risks errors, biases and misunderstandings. Will the AI expert displace the domain expert? Will hypothesis-led research become seen as derivative? Will it lead to a newer version of p-hacking? Already we are seeing more false data and papers, more false citations, more false investigators. It adds to the contamination of the record of science: a record essential to the trusted use of science. The norms and guidelines for the conduct of science and the nature of disclosures will need to change. Exploiting AI cannot be achieved well without appreciating the role of social sciences in ensuring its appropriate use is key.

AI07: Computational Approaches for Small Molecule Drug Discovery.

May L. T. [1]

[1]Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, 3052.

The process of small molecule drug discovery remains an expensive and time-consuming process[1]. Computational methods can predict drug-receptor interactions, accelerating the exploration of chemical space and enabling an improved rate of hit identification. These methods predict drug-target interactions using different types of input data. Sequence-based approaches rely on one-dimensional (1D) inputs, such as ligand SMILES and protein sequences, or two-dimensional (2D) representations, including molecular graphs and predicted contact maps. Structure-based approaches incorporate three-dimensional (3D) protein structures, determined experimentally or predicted *in silico* (e.g., via AlphaFold2). We have employed both structure-based[2,3] and AI-based structure-independent[4] approaches to identify G protein-coupled receptor (GPCR) orthosteric and allosteric small molecule ligands, using the adenosine A_1 receptor (A_1R) as a model system. Structure-based virtual screening was to identify design subtype-selective A_1R antagonists[2], exploiting a non-conserved sub-pocket within the orthosteric binding site[5]. Computational screening of a 4.6 million compound library and subsequent structure-guided optimization yielded subtype-selective A_1R orthosteric antagonists with nanomolar potency. More recently, we identified A_1R positive allosteric modulators (PAMs) using structure-based virtual screening[3]. This virtual screening targeted a challenging membrane-facing extrahelical pocket revealed by a cryo-EM structure[6]. We identified a lead PAM that enhanced A_1R agonist binding and potency, and had improved physicochemical properties relative to known A_1R PAMs.
To complement structure-based efforts, we developed PSICHIC (PhySIcoCHemICal graph neural network) , an artificial intelligence framework that predicts small molecule binding affinity and functional effects directly from sequence data[4]. By learning interpretable interaction fingerprints, PSICHIC achieves state-of-the-art accuracy in virtual screening tasks, even in the absence of structural information. In a screening campaign for A_1R agonists, PSICHIC successfully ranked the sole active novel compound within the top three candidate compounds. The residue- and atom-level interpretability provided by PSICHIC provides insight into the molecular determinants of binding and selectivity.
Together, these approaches illustrate how structural biology and artificial intelligence can be integrated to advance drug development, enabling the discovery of different types of ligands, even for pharmacologically challenging targets.

1. Gaudelet T, *et al*. (2021) Utilizing graph machine learning within drug discovery and development, *Brief Bioinform*, **22**:6, bbab159, doi.org/10.1093/bib/bbab159
2. Matricon P, *et al.* (2023) Structure-based virtual screening discovers potent and selective adenosine A_1 receptor antagonists. *Eur J Med Chem*. **257**:115419. doi: 10.1016/j.ejmech.2023.115419.
3. Nguyen ATN, *et al*. (2025) Structure-based discovery of positive allosteric modulators of the A_1 adenosine receptor. *Proc Natl Acad Sci U S A*. **122**(28):e2421687122. doi: 10.1073/pnas.2421687122.
4. Koh, H.Y. *et al.* (2024) Physicochemical graph neural network for learning protein–ligand interaction fingerprints from sequence data. *Nat Mach Intell* **6**:673–687 doi.org/10.1038/s42256-024-00847-1
5. Glukhova A, *et al*. (2017) Structure of the Adenosine A_1 Receptor Reveals the Basis for Subtype Selectivity. *Cell* **168**(5):867-877.e13. doi: 10.1016/j.cell.2017.01.042.
6. Draper-Joyce, C.J., *et al.* (2021) Positive allosteric mechanisms of adenosine A_1 receptor-mediated analgesia. *Nature* 597:571–576. doi.org/10.1038/s41586-021-03897-2

AI08: Protein function prediction and classification using language models

Natarajan Kannan [1,2]

1. Department of Biochemistry and Molecular Biology, University of Georgia, Athens, USA
2. Institute of Bioinformatics, University of Georgia, Athens, USA

Recent advancements in artificial intelligence and machine learning are revolutionizing various aspects of molecular biology. Notably, models originally designed for natural language processing are making significant impacts on biological sequence analysis, recognizing that biological sequences essentially function as the cell's own ‘language’ for dictating molecular processes. In this presentation, I will provide an overview of protein language models and describe our efforts to harness the power of these models for protein classification and function prediction tasks. Specifically, I will introduce the Phosformer model, an unsupervised deep learning model trained on millions of protein kinase sequences and kinase-peptide pairs to map kinase-substrate interactions at the proteome scale. I will demonstrate the application of these models in predicting and characterizing the cellular substrates of understudied protein kinases and pseudokinases. Additionally, I will discuss ongoing initiatives to apply these models for investigating protein glycan interactions and automated functional annotations.

AI09: Beyond SMILES: A Multimodal Approach for Enhanced Molecular Property Prediction

Nguyen, B.[1]

[1]School of Mathematics and Statistics, Victoria University of Wellington, Wellington, NZ.

Accurate prediction of molecular properties is crucial in drug discovery and chemical biology, yet existing computational methods often rely on limited molecular representations, such as SMILES strings or 2D/3D structural graphs, each with their own shortcomings. We introduce a unified multimodal framework that requires only the SMILES format as input but leverages both predicted 3D conformer structures and advanced contrastive learning techniques to enhance molecular property prediction. Our approach integrates information from the sequential SMILES notation and spatial 3D conformers, capturing a comprehensive range of molecular features relevant to biological activity and chemical reactivity.

The model employs specialized loss functions: InfoNCE for aligning SMILES and conformer embeddings, ConR for regression, and SupCon for classification, alongside feature distribution smoothing (FDS) to address data imbalance, a common challenge in biological datasets. We validate our framework across diverse case studies, including SARS-CoV-2 drug docking score prediction, molecular property benchmarks from MoleculeNet, and kinase inhibitor prediction for JAK-1, JAK-2, and MAPK-14. In all scenarios, our multimodal model consistently outperforms state-of-the-art methods, demonstrating improved accuracy in both regression and classification tasks.

This work underscores the value of integrating multiple molecular representations and contrastive learning to advance predictive modeling in chemistry and biology, with broad applications in drug discovery, molecular analysis, and related fields.

AI10: Chemical space travel connects patients to early stage drug discovery science.

Flanagan J. U.[1,2], Copping J.M.[1], Chand S.[3], Senanayake D.[4], Abbasi H.[3], Lee W-J.[1,2], Dragunow M.[1,2]

[1]Dept of Pharmacology and Clinical Pharmacology, University of Auckland, Auckland, New Zealand. [2]Centre for Brain Research, University of Auckland, Auckland, New Zealand. [3]Auckland Bioengineering Institute, University of Auckland, New Zealand. [4]National eScience Infrastructure, New Zealand.

The chemical space available to support drug discovery is exploding, reaching almost 100 billion readily obtainable compounds. These ultra-large compound libraries go far beyond the scale accessible to traditional wet-lab based high throughput screening. The existence of massive chemical spaces also creates the tantalizing prospect that the specific molecules destined to be our future medicines already exist, all we need to do is find them. Since these ultra-large chemical spaces exist primarily in a digital format, it is only possible to screen them with methods supported by super-computing environments. Inclusion of drug target 3-dimensional atomic structure data means that it is possible to pin-point specific molecules that bind a drug target. Molecular docking is an established virtual screening method, it performs embarrassingly parallel brute-force calculations that create 3D-structures for every compound on the surface of a drug target. Some of largest molecular docking screens are now covering at least a billion molecules. To keep pace with chemical space expansion, AI applications capable of predicting molecular docking outcomes based only on the compound chemistry are being developed, a needed booster for space travel.

So how do we derive benefit in New Zealand from these global advances? our national high performance computing infrastructure creates opportunity to link our drug discovery sciences directly to patient samples giving New Zealand patients opportunity to participate in early drug discovery and potentially identify new chemistry for future medicines. Targeting the underlying neuroinflammation of many brain diseases, we created a precision drug discovery pipeline that seeks to screen ultra-large chemical spaces against new drug targets identifying new chemistry relevant for pre-clinical development. Our molecular docking platform has identified molecules able to modulate inflammation in brain cells. We are now investigating machine learning to increase the scale and speed of our screening platform.

AI11: From wetlands to wildings: the use of deep learning softmax probabilities for environmental monitoring.

Martin, B[1]., Schindler, J. [1], Mason, N.W.H. [1], Shepherd, J.D. [1]

[1]Manaaki Whenua – Landcare Research Group, Bioeconomy Science Institute, Lincoln, NZ.

Deep learning has transformed the way we analyse and process imagery. For environmental monitoring and mapping, encoder-decoder networks, such as U-Net, have revolutionised the generation of maps from remote sensing imagery and LiDAR. We have used U-Net to map land cover for the national Land Cover Database (LCDB) and New Zealand Land Use Carbon Analysis System Land Use Map (LUCAS LUM), among others. In particular, U-Net is able to separate indigenous and exotic forests in Sentinel-2 10m, 10-band satellite imagery with over 90% accuracy.

Classes with lower representation and smaller footprints are much less reliably detected. We investigated whether the softmax probabilities generated by U-Net could be used to detect these more difficult classes, even when the generated probability is very low, for two important but very different environmental domains: wetlands and wilding conifers. For wetlands, we trained models on small but extensively mapped areas of kettle hole ephemeral wetlands in the Central South Island Glacial Geomorphology area (CSIGG), and pastoral wetlands in the Whangapē catchment of the Waikato region, using spring and autumn Sentinel-2 mosaics and monthly "wetness" layers derived from the Sentinel-2 imagery. For wilding conifers, Sentinel-2 mosaics for the period of 2016 – 2022 were used to train a model using the 2016 LUCAS LUM as training labels, and the "exotic forest" class probabilities were then used as a proxy for wilding conifer detection. This domain is also an example of weakly supervised learning, as the labels are inaccurate for wildings.

For both domains, the softmax probabilities proved to be a useful predictor, generating candidate polygons that could be quickly triaged using higher-resolution imagery (e.g. the 0.3m rural area imagery typically collected by regional councils), or via field investigation. Control and regrowth of wildings was also reliably detected. These results expand our ability to efficiently monitor the natural environment.

AI12: Recognising Taonga with AI: Facial Recognition for Kākā Conservation Management

Lensen, A.[1], Maddigan, P.[1], Shaw, R. C.[2], McLeod, T.[3]

[1]School of Engineering and Computer Science, Victoria University of Wellington, Wellington, NZ, [2]School of Biological Sciences, Victoria University of Wellington, Wellington, NZ, [3]Karori Sanctuary Trust (Zealandia), Wellington, NZ.

The re-establishment of kākā (*Nestor meridionalis*) in urban Wellington is a remarkable conservation success, but one that presents new challenges. Most individuals are unbanded, limiting our ability to monitor their movements, survival, and the emergence of threats such as brodifacoum bait consumption. Supported by a $1 million MBIE Smart Ideas grant, our project is developing the world's first AI-based tool for non-invasive individual identification of kākā using facial and beak morphology.

This talk will share early results from the first year of the project. I will detail the development and evaluation of our initial re-identification pipeline, which uses unsupervised and weakly supervised learning to distinguish individuals based on keyframes extracted from fixed-position trail camera footage. The system has shown promising performance in clustering known, banded kākā based on beak features, even under varied lighting and pose conditions. By leveraging multiple keyframes per visit, we improve robustness against occlusion and movement, laying the groundwork for city-scale monitoring.

Our project takes a deliberately interdisciplinary approach, weaving together three strands: artificial intelligence, conservation ecology, and mātauranga Māori. This framing ensures that the tools we develop are technically robust, ecologically grounded, and aligned with Te Tiriti principles. While co-development with iwi will expand in future phases, our foundational approach recognises the significance of kākā as taonga and supports tangata whenua in their roles as kaitiaki.

Together, these early advances demonstrate the feasibility of AI-powered bird identification and the potential for interdisciplinary tools to inform conservation management, strengthen ecological monitoring, and contribute to more responsive, locally grounded environmental decision-making in Aotearoa.

AI13: The role of artificial intelligence in weather system and microclimate research in Antarctica

Katurji, M.[1]

[1]School of Earth and Environment, University of Canterbury, Christchurch, NZ,

Antarctica's extreme and remote environment and the lack of reliable observations presents a hurdle for regional climate modelling verification and detailed microclimate analysis. Historical and future spatial-temporal analysis of near-surface climate and governing weather systems are important for predicting hydrological floods from glaciers, water availability for bacterial life, and understanding fundamental processes governing extreme weather conditions. This talk will explore the advantages of integrating AI into regional and local atmospheric research with particular insights on multi-disciplinary useability in the Dry Valleys of Antarctica.

We will highlight through machine learning techniques how new datasets from satellite remote sensing, numerical weather and climate models, and automatic weather stations can be valuable for our current research problems. Some benefits include improved spatial prediction of extreme weather events at the very local mountainous valley scale and Antarctic wide scale, and the development of high resolution surface climate datasets using integration of machine learning and satellite data. Our upcoming research program in this area will inform future glaciological, hydrological, and ecological research in Antarctica.

AI14: Artificial Intelligence for Aquaculture

Bing Xue[1], Mengjie Zhang[1]

[1]Centre for Data Science and Artificial Intelligence, Victoria University of Wellington, Wellington 6140, New Zealand

Artificial Intelligence (AI) and Machine Learning (ML) have emerged as powerful tools across various scientific domains, revolutionizing data analysis and decision-making processes. These technologies excel in tasks such as prediction, image classification, and pattern recognition, offering unprecedented insights and efficiency. In the realm of aquaculture and marine science, AI and ML applications are rapidly expanding, addressing critical challenges and optimising operations. This talk explores the integration of AI/ML techniques in aquaculture and marine research, focusing on key applications such as mussel farm image analysis and reconstruction, fish breeding optimization, and health monitoring. We will discuss how these technologies enhance mussel farm management, improve yield forecasting, and contribute to sustainable aquaculture practices. By leveraging these advanced computational methods, researchers and industry professionals can drive innovation and sustainability in aquaculture and marine science.

AI15: Machine learning to identify genetic growth variants in snapper

Blommaert, J.[1], Vander Velpen, S.[1,2], Bayer, P.[3], Catanach, A.[1], Jesson, L.[1], Wellenreuther, M.[1,4]

[1] The New Zealand Institute for Plant and Food Research Limited, Nelson Research Centre, Nelson, New Zealand,[2] Hogeschool Leiden, Leiden, The Netherlands, [3] Minderoo Foundation, Perth, Australia, [4] The School of Biological Sciences, The University of Auckland, New Zealand

To build resilience in aquaculture and to meet the growing global food demands, species diversification and genetic improvement are essential. The Australasian snapper (*Chrysophrys auratus*) is a promising warm water-adapted candidate species, supported by a selective breeding programme at Plant & Food Research for over two decades, with genomic selection incorporated in the past decade. In this study, we combined phenotypic and genomic data from the 4th generation snapper cohort to identify genetic variants associated with growth-related traits using a multi-pronged approach: genome-wide association studies (GWAS), machine learning (ML) with XGBoost, and a custom statistical method for structural variants (SVs).

High-throughput, image-based phenotyping was used to quantify 13 morphological traits alongside manual measurements of weight and fork length. Heritabilities were estimated, and trait correlations explored. GWAS identified 24 single nucleotide polymorphisms (SNPs) significantly associated with growth, many of which mapped to genes involved in metabolic and developmental pathways. ML models trained on SNP genotypes achieved moderate predictive power, with notable overlap between SNPs identified by GWAS and those prioritised by ML.

We further incorporated SVs, which are known to influence complex traits, into our analyses. Our custom approach centred around calculating average phenotypes per SV depending on their presences in each sample. Through this, we detected more growth-associated SVs than either ML or GWAS, with many SVs located near genes with growth-relevant functions. These findings underscore the importance of including SVs in genomic prediction models and breeding decisions. However, challenges such as sequencing costs and limited sample sizes constrain the power of both GWAS and ML methods for SV detection.

Together, our findings support the integration of computer vision-based phenotyping, GWAS, ML, and SV analysis to improve trait prediction and genomic selection in aquaculture breeding programmes.

AI16: TaxonGPT: Taxonomic Classification Using Generative Artificial Intelligence

Haoyuan Huang[1], Teng Li[1], Zhixuan Wang[1], David Seldon[1], and Allen Rodrigo[1]

[1]The School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand

Although biological taxonomy has well-established workflows, it is estimated that only 10% of species have been fully named or described taxonomically. La Salle et al[1] discuss what they refer to as the "Taxonomic impediment": uneven distribution of global taxonomic resources, lack of unified data standards, and over-reliance on taxonomists' expertise. Specifically, taxonomists in some regions lack sufficient resources to conduct research, and the absence of consistent data standards hinders the aggregation and sharing of taxonomic data.

To address the challenges of consistency and quality in taxonomic classifications, we have developed a Python program called TaxonGPT, that utilizes the natural language processing capabilities of generative artificial intelligence (gAI, specifically, ChatGPT-4o) to generate taxonomic descriptions and taxonomic keys. To counter the propensity for large language model gAIs to "hallucinate", we use knowledge graph semantic representation and an error-checking module to ensure that accurate taxonomic descriptions and keys are obtained. In this presentation, we describe how TaxonGPT embeds ChatGPT-4o's responses as outputs. We also report on benchmark tests for accuracy, efficiency and reproducibility. Our results show that for all simulated (n = 50) and real (n = 11) datasets we used, TaxonGPT generated descriptions and taxonomic keys without any errors.

We conclude our presentation with a look at some further work, including the development of a pipeline that includes phylogenetic reconstruction. Finally, we discuss the current limitations of our work, and opportunities for the future use of gAIs in similar biological applications.

[1]La Salle, J., Wheeler, Q., Jackway, P., Winterton, S., Hobern, D., & Lovell, D. (2009). *Accelerating taxonomic discovery through automated character extraction.* Zootaxa 2217:43–55.
[2]Kim, K. C., & Byrne, L. B. (2006). *Biodiversity loss and the taxonomic bottleneck: emerging biodiversity science.* Ecological Research, 21:794–810.

AI17: When AI meets two outlaws: Exploring interactions in biological systems, non-invasively.

Reis, M.M.[1]

[1]Bioeconomy Science Institute

The interactions among molecular species and structures play a critical role in biological processes and functions. However, probing these interactions is not straightforward, as it requires non-invasive methods that do not disrupt the systems being studied. Spectroscopy-based techniques can be non-invasive and provide insights into molecular composition and structure, but their signals are often complex and not easily interpretable.

Spectra in the Near-Infrared (NIR) range result from interactions among chemical species and can reveal rich, in-depth information about biological samples. These spectra can be acquired using hyperspectral imaging (HSI) systems, which generate spatially resolved maps of chemical and structural information.

While HSI offers a window into interaction networks within biological systems, extracting meaningful information from the complex spectral and spatial data remains a challenge. This is where artificial intelligence (AI), including machine learning, become crucial, enabling the translation of data complexity into actionable insights. Our hypothesis is that hyperspectral imaging in the NIR region provides access to the "interactome" of biological systems.

In a series of case studies, we identified supporting evidence for this hypothesis. For instance, bacterial spores can survive heat treatment and become active under specific biological triggers. Our research shows that combining hyperspectral imaging with advanced analytics allows us to detect spores and determine whether they are live, dead, active, or dormant. Similarly, we found that assessing digestion, a dynamic process involving chemical and structural changes, benefits from analyzing spatial information to reveal previously unseen interactions.

Emerging AI models are poised to further enhance our understanding of these complex molecular interactions. In this presentation, I will explore how hyperspectral imaging, molecular interactions, and system-level insights can be aligned with AI-based approaches to advance biological understanding of interactions in biological systems.

Acknowledgement: Funding from the Food Integrity programme (Strategic Science Investment – AgResearch).

AI18: Designing for Dependability: Embedding Reliability and Robustness in AI Systems

Chang, X.[1], Zhao, K.[1], Dost, K.[2], Dobbie, G.C.[1], Wicker, J.S.[1]

[1]School of Computer Science, University of Auckland, NZ, [2]Jožef Stefen Institute, Ljubljana, Slovenia.

As AI systems increasingly mediate critical decisions across healthcare, finance, and public services, their reliability and robustness have become central ethical and technical challenges. Current deployments frequently reveal fragility in real-world conditions, manifesting as unexpected failures, degraded performance on out-of-distribution data, and susceptibility to adversarial manipulation. This paper addresses the problem of ensuring that AI systems operate consistently and safely under diverse, unforeseen circumstances.

This presentation explores how reliability and robustness can be tested for in the design and upkeep of AI systems, drawing lessons from traditional software engineering. Bridging technical foundations with applied considerations, the talk offers practical insights into building AI systems that are not only effective but also resilient and ethically sound.

Our research highlights that reliability and robustness are not merely technical optimisations but foundational to trustworthy AI deployment. Embedding these principles from the design stage through to post-deployment monitoring significantly mitigates risks of harm and supports the development of AI systems that are both effective and dependable in real-world environments.

AI19: AI Research in Health NZ | Te Whatu Ora

Whittaker, R[12], Dobson, R.[12]

[1]AI Lab, Health NZ, [2]TRANSFORM, University of Auckland.

Artificial Intelligence (AI) tools are in use in many health services globally. Some forms of AI are still relatively new with little evidence base behind their accuracy, effectiveness, safety, bias and unintended consequences when implemented in real world contexts. The promise of improved technical accuracy, such as with computer vision and diagnostic imaging, and of system efficiencies with reductions in administrative and documentation tasks, is creating huge pressure to implement more rapidly than previous innovations. Also, AI tools are currently unregulated in New Zealand - all leading to the potential for harm to occur if implementation of AI tools is not supported by rigorous assessment and ongoing research and monitoring.

For these reasons, Health NZ | Te Whatu Ora has established a National AI & Algorithm Expert Advisory Group (NAIAEAG) to review proposals from clinicians and health services who wish to develop and/or implement AI tools. This group of multidisciplinary internal and external experts and representatives have developed a checklist for the review and assessment of AI tools.[1] The group is also working on refining this with a new evaluation framework in collaboration with New Zealand AI academics.

The NAIEAG is supported by an internal AI Lab that has several functions, including testing externally developed AI tools within the Health NZ data environment, developing simple tools with our population datasets, and conducting research and evaluation on the implementation of AI tools within New Zealand health services.

This presentation will outline some of the priority areas for AI research within Health NZ. This should help other researchers to understand the potential and the pitfalls, and also the relevant processes they need to be aware of to conduct high quality AI health research in collaboration with New Zealand's public health services.

1. Whittaker R, Dobson R, Jin CK, Style R, Jayathissa P, Hiini K, Ross K, Kawamura K, Muir P; Waitematā AI Governance Group. An example of governance for AI in health services from Aotearoa New Zealand. NPJ Digit Med. 2023 Sep 1;6(1):164. doi: 10.1038/s41746-023-00882-z.

AI20: Artificial Intelligence and public health policy: a measles modelling case study

Callaghan, F.M.[1], Orsi, A.[2]

[1]Public Health Agency, Ministry of Health, Wellington, NZ, [2]PHF Science, Kenepuru, NZ

Artificial Intelligence (AI) tools are having a significant impact in the healthcare space. Following the Precision Health Insights Briefing published by the Ministry of Health, the Prime Ministers Chief Science Advisors report on AI in healthcare, and Public Service AI Framework, the Ministry of Health has developed a work program for precision health and AI tools. In this work, we give the current overview of the AI policy landscape and the Ministry of Health's roadmap for future work in AI, including establishing governance groups, policy options, social licence, ethics, and research needs.

Furthermore, we illustrate some of the current challenges and opportunities with a case study of measles modelling using a synthetic population and overlaying novel AI modelling techniques through the ALMA Digital Twin platform. This platform implements a novel Agent-based modelling technique to study the resilience of Aotearoa New Zealand to a Measles outbreak across regions. We also explored the impact of different interventions in terms of different vaccination rates across demographics. In this context, the use of synthetic data helped to address many of the data privacy issues, while the AI model was able to estimate several epidemic trajectories and provide insight on vaccination strategies.

1. Ministry of Health, Precision health: AI and genomics website: https://www.health.govt.nz/strategies-initiatives/programmes-and-initiatives/precision-health-ai-and-genomics [accessed 08 July 2025].

AI21: Machine Learning Unravel Gene Signatures of Coronary Artery Disease Comorbidity with Periodontitis

Liu, W.T.[1], Yang, N.N.[2], Yang, W.Q.[2], Wang, S.B.[2], Jin, Y.[3], Wang, K.[3], Tan, L.G.[3], Zhang, H.[3], Huang, Z.[3], Chen, Q.[3], Xing, H.H.[2]

[1]School of Biological Sciences, University of Canterbury, Christchurch, NZ, [2] Healthcare Big Data Center, School of Public Health and Management, Hubei University of Medicine, Shiyan, Hubei, China, [3] Renmin Hospital, Affiliated Hospital of Hubei University of Medicine, Shiyan, Hubei, China.

Emerging evidence suggests a complex interplay between periodontal disease (PD) and coronary artery disease (CAD) risk[1-2]. The novel insights into the shared pathogenesis of PD and CAD will potentially inform future therapeutic strategies. We employed machine learning methods to explore gene signatures implicated in the progression of PD to CAD.

Firstly, differentially expressed genes (DEGs) in co-expressed gene modules by weighted gene co-expression network analysis (WGCNA) for PD and CAD were explored from GSE10334 and GSE66360, respectively. The common 48 co-expressed DEGs for PD and CAD were selected as candidate gene signatures. Four machine learning algorithms, SVM, RF, XGB, and GLM, were employed to construct predictive models, and finally five gene signatures were selected using external PD dataset GSE106090 and CAD dataset GSE179789 as validation cohort. CIBERSORT (Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts) algorithms indicate the five gene signatures highly correlated with eosinophils, activated memory CD4+ T cells, and resting dendritic cells. Moreover, RNA-sequencing on blood samples of five patients with both CAD and PD compared five healthy controls validate four markers of them, *FOS, MME, PECAM1,* and *VNN2*, suggesting their potential as key targets for future investigations.

Compared with statistical differential test, machine learning can unravel the complex hidden data patterns. This study shows that the gene signatures confirmed from various machine learning algorithms make more biological sense and robustness, suggesting the power of AI on biomarker detection.

References

1. Shetty B, Fazal I, Khan SF, Nambiar M, D KI, Prasad R, Raj A (2023). *Association between cardiovascular diseases and periodontal disease: more than what meets the eye.* Drug Target Insights.17:31-38.
2. Sumayin Ngamdu K, Mallawaarachchi I, Dunipace EA, Chuang LH, Jafri SH, Shah NR, Jeong YN, Morrison AR, Bhatt DL (2022). *Association Between Periodontal Disease and Cardiovascular Disease (from the NHANES)*. Am J Cardiol.178:163-168.

AI22: AI-Based Drug-Drug Interaction Prediction: A Scoping Review of Model Robustness, Clinical Relevance, and Implementation Challenges

Reta D[1], Mohammed M[1], Martini N[1]

[1] School of Pharmacy, Faculty of Medicine and Health Sciences, the University of Auckland, NZ.

Background: Globally, medication errors are estimated to cost over USD 42 billion annually, with New Zealand incurring over $280 million in three years. Drug-drug interactions (DDIs) are a significant contributor to medication-related harm, particularly in the rising of polypharmacy. Clinically evident DDIs are reported by clinicians in up to 17.2% of patients, while potential DDIs identified through electronic interaction checkers can reach 64.9%. Although only a small portion of these leads to documented harm (2.2% of clinically evident and 1.3% of potential DDIs), the sheer volume of possible DDI pairs exceeds a clinician's capacity for recall. Existing rule-based clinical decision support systems (CDSSs) are limited by high false-positive rates, alert fatigue, and override rates exceeding 90%. Artificial intelligence (AI) offers a promising solution to enhance predictive accuracy and reduce cognitive burden.

Objective: To map the landscape of AI-based approaches for DDI prediction, examining model types, data sources, features, validation strategies, metrics, challenges, and clinical relevance.

Methods: A scoping review of studies published in English between 2020 and 2024 was conducted across five databases, guided by PRISMA Extension for Scoping Reviews (PRISMA-ScR).

Results: Of 823 studies, 112 from 22 countries were included. Notably, deep learning (n=67, 59.8%), supervised learning (n=99, 88.4%), and internal validation (n=91, 81.3%) dominated, with limited external real-world validation. DrugBank-like structured databases were the primary data sources, where chemical descriptors served as the key input features, and only eighteen studies—mostly deep learning—achieved ≥90% across key performance metrics. Only 17 (15.2%) studies incorporated patient-related input features. Substantial limitations include a lack of patient-specific context, a binary classification focus, limited generalizability, and poor interpretability.

Conclusions: Despite challenges, AI—particularly deep learning—shows promise in enhancing DDI prediction. Clinically validated domain-specific AI models co-designed with healthcare professionals that address DDI severity level may improve trust, interpretability, and clinical relevancy.

AI23: Rethinking refeeding: Can the neurocognitive developmental delays associated with childhood undernutrition be reversed?

Portlock, T.J. [1,*], Shama, T.[2,*], Kakon, S.H.[2,*], Pook, C.[1,*], Gluckman, P.[1], Forrester, T.[7], Haque, R.[2], Nelson, C.A.[3], O'Sullivan, J.[1456]

[1] The Liggins Institute, University of Auckland, NZ, [2] Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh, [3] Department of Pediatrics, Boston Children's Hospital and Harvard Medical School; Harvard Graduate School of Education, Boston, USA, [4] The Maurice Wilkins Centre, The University of Auckland, New Zealand, [5] MRC Lifecourse Epidemiology Unit, University of Southampton, UK, [6] Singapore Institute for Clinical Sciences, A*STAR, Singapore, [7] Faculty of Medical Sciences, UWI Solutions for Developing Countries, The University of the West Indies, Jamaica

* These authors contributed equally

Background: Malnutrition contributes to nearly half of all childhood deaths globally. Current treatments focus on rapid weight gain, often neglecting long-term cognitive recovery[1].

Objective: We conducted a randomized controlled trial in Bangladesh to evaluate whether a brain-targeted, micronutrient-enhanced supplementary food (E-RUSF, Nutriset, France) could improve cognitive outcomes in one-year-old children with moderate acute malnutrition (MAM), compared to locally produced Ready to Use Supplementary Food (RUSF) and well-nourished controls.

Methods: Participants were assessed for physical recovery, cognitive and executive function, brain activity (EEG), gut microbiome composition, plasma metabolomics, and polygenic scores for educational attainment. Interpretable machine learning approaches were used to uncover predictors of recovery and brain outcomes.

Results: While the enhanced nutritional intervention accelerated weight gain, it did not improve recovery rates or cognitive performance relative to local RUSF. Recovery was most strongly associated with pre-existing factors including head circumference, weight, and genetic predisposition. Multi-omics integration revealed a mechanistic link between gut microbial fermentation of dietary proline and improved brain activity and vocalisation behaviour, highlighting the role of nutrient uptake efficiency over composition alone.

Conclusion: Our findings indicate that neurodevelopmental recovery following MAM is shaped by complex interactions between diet, microbiota, metabolism, and host factors, and is not significantly improved by nutritional supplementation alone. These results suggest that early-life social, biological, and genetic conditions constrain the potential for recovery and should be considered in the design of future interventions. This work underscores the need to expand outcome measures in nutritional rehabilitation beyond anthropometry to include brain and behavioural endpoints.

1. Portlock T., Shama T., Kakon S.H., et al. (2025). *Interconnected pathways link faecal microbiota, plasma lipids and brain activity to childhood malnutrition-related cognition*. Nature Communications. 16: 473. https://doi.org/10.1038/s41467-024-55798-3

AI24: Using machine learning with routinely collected health data for identification of dementia risk in Aotearoa New Zealand.

Gonzalez-Prieto, C.,[1] Dobbie, G.,[1] Wilson, D.,[1] Rivera-Rodriguez, C.,[2] Yates, S.,[3] & Cullum, S.[3]

[1]School of Computer Science, University of Auckland, NZ, [2]Department of Statistics, University of Auckland, NZ, [3]Department of Psychological Medicine, University of Auckland, NZ

Background: There are an estimated 80,000 people living with dementia in New Zealand (NZ), with greater risk in Māori and Pacific Islanders due to higher prevalence of risk factors (e.g. diabetes, obesity and hypertension). Only half of dementia is diagnosed due to a lack of skilled clinicians. This is a global problem addressed by several research groups worldwide with successful use of artificial intelligence (AI) and health data to identify dementia, but most have been conducted in highly curated research samples. Our research group aimed to deploy machine learning with routinely collected health data to demonstrate effective dementia identification in a 'real-world' clinical setting.

Methods: Routinely collected health data comprising sociodemographic and clinical data from the Counties Manukau 65+ population were extracted and deidentified by the Health Informatics Department, Middlemore Hospital. Dementia status was ascertained via pharmaceutical records (antidementia prescriptions), interRAI assessments (dementia diagnoses), and ICD-10 diagnostic codes. A nested case-control design was implemented, using pre-diagnostic information (including pharmaceutical prescriptions, blood test results, comorbidities and delirium screening) for prediction. Machine learning models capable of handling both longitudinal and cross-sectional data, including Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), were assessed using a training/testing framework. Data processing was conducted in R and machine learning modelling in Python.

Results: The analysis was conducted on a sample of 8195 cases and 8195 age and data-length matched controls. The RNN model demonstrated accuracy of 78.6% in identifying dementia cross-sectionally (sensitivity:71%, specificity:87%). The model also achieved an accuracy of 60.3% one year before diagnosis (sensitivity:35%, specificity:85%) and an accuracy of 57.3% eight years before diagnosis (sensitivity:26%, specificity:88%).

Conclusion: Machine learning models and routinely collected health data may be a potent tool for dementia prediction in both NZ and globally, offering the potential for early intervention to prevent or delay the onset of dementia.

AI25: AI-based implementation of NMR-guided potentials for dramatically coarse-grained protein simulations.

Cullen M. S.[1], Biancaniello C.[2], Taskova K.[3], Miletic V.[4], De Simone A.[2], Mercadante D.[1]

[1]School of Chemical Sciences, University of Auckland, Auckland, NZ, [2]Department of Pharmacy, University Federico II, Naples, Italy, [3]School of Computer Science, University of Auckland, Auckland, NZ, [4]Max Planck Computing and Data Facility, Munich, Germany.

Molecular dynamics is a functional feature of proteins. The paradigm of a dynamics-driven protein function is shown at its extremes by intrinsically disordered proteins (IDPs), which despite lacking well-defined structure, carry complex tasks inside cells.[1] Driven by extreme plasticity, IDPs quickly interconvert between conformations, making them challenging to study experimentally. All-atom molecular dynamics (MD) simulations can provide detailed conformational ensembles but also suffer severe limitations, since IDPs populate extended conformations requiring large computational power in order to be simulated exhaustively.

Dramatically coarse-grained (CG) modelling of IDPs, with each amino-acid mapped to a single bead, have tackled this challenge and improved on our capacity to simulate IDPs.[2] Nonetheless, these models blatantly ignore the formation of transient secondary structure (SS), which is crucial for IDPs' function.

Following on previous work[3], here we trained Napshift-Cα, a neural network (NN) that can predict chemical shifts (CS) from Nuclear Magnetic Resonance (NMR) experiments for single bead-per-amino-acid CG models. We show how an AI-enhanced mapping between atomistic NMR observables and coarse-grained representations of proteins can guide CG simulations to sample larger systems for longer timescales, while taking into account the secondary structure propensity of IDPs natively proxied by CS. To integrate CS into MD simulations, we developed a plugin that, using the trained NN, predicts NMR CS on the fly and uses the difference between predicted and experimental CS as a restraint to reproduce experimentally sound NMR-guided ensembles. To the best of our knowledge, this approach represents the first direct integration of NMR data into severely coarse-grained MD simulations.

References

1. Babu, M.M. *Versatility from protein disorder*. *Science (1979)* **337**, 1460–1461 (2012).
2. Tesei, G. *Conformational ensembles of the human intrinsically disordered proteome*. *Nature* **626**, 897–904 (2024).
3. Qi, G. *Enhancing Biomolecular Simulations with Hybrid Potentials Incorporating NMR Data*. *J Chem Theory Comput* **18**, 7733–7750 (2022).

AI26: Bridging Clinical and Consumer Electroencephalography: A Transfer Learning Approach for Artifact Detection

Barnes L.D.[1], Vrijdag X.C.E.[1], Hallum L.E.[2]

[1]Department of Anaesthesiology, University of Auckland, NZ, [2]Department of Mechanical and Mechatronics Engineering, University of Auckland, NZ.

Machine-learning (ML) can enhance electroencephalography (EEG) diagnostics, but ML requires large, expert-annotated datasets. We wondered whether training a model on a large, expert-annotated dataset enabled use on data from a consumer-grade EEG device (Muse 2; InteraXon, Toronto).

We developed a transfer learning framework using clinical and consumer-grade EEG to train a binary classifier to detect eye-blink artifacts. Our model followed the EEGNet architecture, a three-layer convolutional neural network. We first evaluated the model on Muse data, then re-evaluated it after pre-training on clinical data. This clinical EEG, sourced from the Open Science Foundation (OSF), varied in channel configuration and sampling rate. We re-referenced to the average, re-sampled to 256 Hz, and selected channels corresponding to the Muse headset (AF7, AF8, TP9, TP10). In both datasets, participants fixated a visual stimulus, which was used to cue eye blinks. OSF segments were labelled using four-channel electrooculography; an EEG expert manually annotated the Muse data.

To assess pre-training benefit, we compared a model trained on Muse data to one pre-trained on OSF and fine-tuned with Muse data (using 8 second training segments). Using leave-one-participant-out cross-validation on the full Muse dataset (191 minutes across 10 participants), the model achieved a Matthews correlation coefficient (MCC) of 0.34, which was statistically significantly greater than chance (i.e., MCC = 0.0). However, when the model was pre-trained on the OSF dataset (228 minutes of data across 50 participants) and re-evaluated on the Muse dataset, it achieved 50%-maximum performance (i.e., MCC = 0.17) in a fraction of the time (8.4 vs 29.4 minutes, Student's t-test, $t = 68.5$, $p < 0.001$).

This study illustrates how transfer learning can bridge the gap between clinical datasets and consumer hardware. Although validated for artifact detection, the approach may generalize to broader EEG applications, supporting scalable, low-cost neurodiagnostic tools beyond traditional clinical environments.

AI27: Real-Time, Label-Free Assessment of hPSC Quality Using AI-Powered Cell Counting and Colony Morphology Analysis

Britney Ragunton, Steve Van Buskirk, Devin Wakefield, Ninad Ranadive, Andrew Pipathsouk, Baikang Pei, Hong Zhou, Tracy Yamawaki, Mike Berke, Chi-Ming Li, Christopher Hale, Songli Wang, and Stuart M. Chambers.

Human pluripotent stem cells (hPSCs) are foundational to many applications in regenerative medicine, disease modeling, and cell therapy, but their successful use at clinical and industrial scale depends on reproducible and scalable quality control. Traditional assays for assessing hPSC quality, including immunocytochemistry, flow cytometry, and manual microscopy, are destructive, time-consuming, and poorly suited for continuous, high-throughput workflows. To overcome these limitations, we developed a label-free, automated imaging platform that uses artificial intelligence (AI) to perform real-time, quantitative analysis of hPSC cultures using live-cell, phase contrast microscopy.

The system consists of two deep learning models that operate in tandem. The first is a cell counting model that estimates the number of nuclei in each brightfield image, enabling a label-free proxy for cell density and proliferation dynamics. This model uses a convolutional neural network trained to predict spatial density maps, where the integral over the map yields the estimated number of cells. Ground truth for training was obtained by co-registering phase contrast images with fluorescence images stained for nuclear markers (e.g., Hoechst), allowing pixel-level supervision of nuclear density. The model learns to recognize and encode subtle textural and spatial features associated with nuclear presence, even in densely packed colonies where nuclei are not individually resolvable.

This density map–based approach allows for robust estimation of local cell density, capturing spatial heterogeneity within and between colonies. The model achieved a mean absolute error of less than 5% when validated against held-out fluorescence-labeled datasets and generalized well across different hPSC lines and seeding densities. By enabling longitudinal tracking of cell counts in a completely non-invasive manner, this model serves as a foundational layer for automated culture monitoring, supporting decisions such as media changes, passaging, or identifying abnormal growth patterns.

Building on this, the second AI model focuses on colony segmentation and pluripotency assessment. A convolutional neural network segments colonies and extracts quantitative morphological features including area, eccentricity, solidity, boundary roughness, and texture entropy. These features are embedded into a latent representation space that is used to infer a pluripotency score, which was benchmarked against immunocytochemistry for OCT4 and NANOG. Morphological changes indicative of spontaneous differentiation, such as loss of colony circularity and increased edge irregularity, were detectable up to 48 hours before fluorescence markers indicated loss of pluripotency. The inferred scores correlated strongly with expression-based ground truth (Pearson $r > 0.9$).

Together, these models provide a comprehensive, real-time picture of hPSC culture status from label-free images alone. The platform operates within standard incubators, supports continuous time-lapse imaging, and enables automated, non-destructive quality control at colony and population levels. This AI-driven approach eliminates the need for manual annotation or destructive endpoint assays and represents a significant advance toward closed-loop control of hPSC manufacturing processes. Future work will focus on

incorporating lineage-specific differentiation classifiers and interfacing the system with robotic culture platforms for fully autonomous stem cell production.

Summary of Abstracts for the Poster Session Template

No.	Title	Presenter	Institutions
P61	AI-Powered detection of growth and division in fish cell lines	Jessica Ispada	Plant & Food Research Group, Bioeconomy Science Institute, Nelson Research Centre, Nelson, NEW ZEALAND
P62	DNAmBERT: A Transformer-Based Model for Non-Invasive Cancer Detection Using DNA Sequence and Methylation Data	Maryam Yassi	University of Otago, Dunedin, NEW ZEALAND University of Auckland, Auckland, NEW ZEALAND UPES University, Dehradun, INDIA

P61: AI-Powered detection of growth and division in fish cell lines

Ispada, J.[1], Algie, M.[1], Ashton, D.T.[1], de Vries, I.[1], Böhmert, B.[1], Chong, G.L.W.[1], Dowd, G.C.[1]

[1] Plant & Food Research Group, Bioeconomy Science Institute, Nelson Research Centre, Nelson, NZ.

Manual quantification of cell growth and division is time-consuming. To address this, artificial intelligence (AI) has been successfully applied to automate image analysis across various cell types. In this study, we trained an AI model using two fish cell lines of aquacultural relevance - CAtmus1PFR and OTgill1PFR - which exhibit distinct cell and nuclear morphologies. Cells were stained with Syto-9, a nucleic acid fluorescent stain for 20 minutes. Images were captured at 10X and 20X magnification using the IncuCyte® Live-Cell Analysis Systems. A total of 85 images were randomly selected and used on Roboflow for model training (YOLOv8). Nuclei were labelled as either “nuclei,” “met” (metaphase/early mitosis), or “ana” (anaphase or telophase/late mitosis). The model was trained with the following parameters: resizing to 640×640 pixels, horizontal/vertical flipping, 90° rotations, and minor adjustments to hue, saturation, brightness, exposure, and blur. This yielded 447 augmented images used for training. Model evaluation showed 81.0% of recall and 88.0% precision, with mean average precision (mAP50) ranging from 83–89% for nuclei, 14–40% for meta, and 35–37% for ana for validation and test datasets. The model was deployed to a cloud-hosted application Morphometrics 2.0, a Python-based user-friendly application, to process batches of 100 images (~8 minutes), generating CSV outputs and labelled images. The model was used to analyse 1,920 images in under 3 hours as part of a 3×2×2 experimental design collected on multiple time points to estimate cell growth. The automated labelling yielding 424,212 nuclei, 2,867 meta, and 915 ana counts in total. Visual inspection confirmed high accuracy, with missed detections primarily due to low fluorescence intensity or high background noise in some image sets. Mislabelled images will be used to retrain the model after manual labelling. The current system enables efficient growth curve estimation and condition-based comparisons at scale.

P62: DNAmBERT: A Transformer-Based Model for Non-Invasive Cancer Detection Using DNA Sequence and Methylation Data

Maryam Yassi [1,2], Mark Ezegbogu [2], Peter Stockwell [2], Ben Brockway[3], Rajiv Kumar [3], Euan J. Rodger [2], Aniruddha Chatterjee [*2,4], Matthew Parry [*1,5]

[1] Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand. [2] Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand.[3] Department of Medicine, Dunedin School of Medicine, University of Otago, New Zealand.[4] Honorary Professor, UPES University, Dehradun, India.[5] Te Pūnaha Matatini Centre of Research Excellence, University of Auckland, Auckland, New Zealand.

*Joint senior authors.

Abstract:

***Background*:** DNA methylation alterations are a hallmark of cancer and provide promising biomarkers for non-invasive diagnosis using circulating cell-free DNA (cfDNA). However, existing computational models often fail to fully integrate the complex contextual relationships between DNA sequence and methylation patterns and rich information available at a sequence read level.

***Aim*:** This study presents DNAmBERT, a novel deep learning model based on the Transformer architecture, designed to integrate DNA sequence and methylation data for accurate, robust, and scalable non-invasive cancer detection.

***Methods*:** DNAmBERT uses a k-mer tokenization of DNA reference sequences combined with whole-read methylation haplotypes encoded using a hybrid input format separated by a [SEP] token. The model is pre-trained using a masked language modelling and fine-tuned with attention-based pooling for classification tasks. We evaluated the model performance on cfDNA methylation datasets across various sequencing platforms and cancer types.

***Results*:** DNAmBERT achieves high classification performance AUC > 0.95 in binary cancer detection across colorectal and lung cancers. The model is further extended to multi-cancer classification (colorectal, lung, and liver cancers), achieving an average AUC of 0.96, demonstrating strong generalization and interpretability. Transfer learning experiments show that DNAmBERT adapts effectively to underrepresented cancer types and early-stage samples, even under imbalanced data conditions.

***Conclusion*:** By integrating DNA sequence and methylation information in a Transformer-based framework, DNAmBERT provides a powerful tool for non-invasive cancer detection using cfDNA. Its performance across multiple cancer types and sequencing platforms highlights its potential for clinical translation and large-scale screening applications.