BOARD # 223: AI-UPP IRES Year 1: Program Development and Initial Lessons Learned

Prof. Mark A Chapman, University of San Diego

Mark Chapman is an associate professor at the University of San Diego in the Department of Integrated Engineering. His interests lie in the fields of skeletal muscle mechanics, muscle disease, exercise physiology, international education and engineering education. He earned his MS and PhD in bioengineering from the University of California, San Diego and a B.S. in biomedical engineering from the University of Minnesota.

AI-UPP IRES Year 1: Program Development and Initial Lessons Learned

This International Research Experience for Students (IRES) NSF project (IRES: Artificial Intelligence and Data Science for the Understanding, Prediction and Prevention of Disease (AI-UPP)) focuses on creating an immersive international summer research experience in Stockholm for students enrolled in a primarily undergraduate institution (PUI). Each year, eight undergraduate students enrolled in bioengineering, biomedical engineering, computer science, bioinformatics and/or computational/mathematical biology programs are sent to Stockholm for 10 weeks to work with one of seven host labs at the Karolinska Institute (KI). The students work on projects which have specifically been co-designed by host lab researchers and the program director to provide students with a structured research project at the appropriate academic level which also meaningfully contributes to the host lab's research program. In addition, the students participate in pre-departure research training, pre-departure cultural training, professional development workshops, a visit to the European Centre for Disease Prevention and Control (ECDC), an open forum on the use of AI in health, cultural outings in Stockholm, a final research presentation symposium, a career readiness workshop series, and be able to present their work at the Biomedical Engineering Society (BMES) annual meeting. This paper reports on the first year of the grant, program structure and discusses the efforts by the program director (PI) to properly prepare students to conduct their summer research project.

Overarching Theme and Research Focus of IRES Site

Over the past few decades the generation of 'big data' within biology and health has exploded. Within biology, this is largely due to technological improvements within the '-omics' space which has allowed for the collection of massive amounts of genomic, transcriptomic, proteomic and metabolomic data with applications from basic science to medicine [1]. Recently, the advent of single-cell and spatially-resolved omics techniques has further contributed to the large amounts of biomedical data being collected and analyzed every year [2]. Apart from these -omics data types, big data within biology and health also relates to studies involving high-throughput microscopy and medical imaging [3]. With ever-improving microscopy and medical imaging techniques, it is becoming easier to amass large amounts of raw imaging data. With the availability of these large datasets, new challenges are arising related to the analysis and interpretation of these data. As such, we turn to AI and data science techniques to efficiently, accurately and precisely mine data to maximize our understanding of various biological systems [2],[4]. Although significant advances have been made in this area, new AI and data science methods are needed as existing biotechnologies improve and new methods emerge which will increase the volume and complexity of these datasets.

Data analytics as applied to large biological data presents unique challenges related to student training. Students graduating with a bachelor's degree in biology traditionally lack the computational skills necessary to use and/or develop big data analysis tools to maximize the amount of information gained from a particular experiment. However, while students with degrees in computer science are able to develop computational tools to analyze biological datasets, these students can lack the biological knowledge needed to design the correct experiments and/or interpret biological significance of their findings. The fields of bioengineering, biomedical engineering, bioinformatics, and computational/mathematical biology attempt to bridge this disconnect by providing students with training in both computational methods and the biological sciences. While these interdisciplinary fields provide a foundation in computer science and biology, students are often in need of co-curricular training activities, such as academic research, to fully integrate the disparate fields within their degree

program. As such, this IRES program provides a valuable research experience to train students interested in AI and data science as applied to the biological sciences.

A main aim of this program is to provide research experiences for individuals who are historically underrepresented in engineering/computer science research. This program heavily focuses on the recruitment, training and mentorship of underrepresented groups. Improving diversity in STEM is critical to ensuring an equitable future, especially in the context of this proposal since biases are known to exist within AI algorithms as applied to health.[5] Looking to the future, as the shortage of physicians worsens, AI will be used to expedite patient screening, which has been shown to have significant issues with algorithm bias. For example, AI systems for computer-assisted diagnosis have been shown to have an issue with 'underdiagnosis bias' in under-served patient populations[6] and when analyzing datasets with gender imbalances[7]. If these biases remain unchecked in the future, they could worsen existing health disparities. These well-documented cases of algorithm bias typically stem from a lack of diversity in the test datasets used to train an algorithm, a lack of diversity on AI design teams, or a combination of these two factors [8]. Without a concerted effort to increase diversity in the fields of AI and data science, there is a risk of further exacerbating these biases embedded in AI technologies. Simply diversifying the design team is only one way to address algorithm bias, which is why this program also trains students in the concept of algorithm bias.

Rationale for Conducting the Project Internationally

The research context of this IRES site is strongly rooted in Sweden's social and political culture. Sweden has one of the oldest universal healthcare systems in the world and it stems from the Swedish political concept of *Folkhemmet*, or the people's home [9]. *Folkhemmet* is essentially the concept that everyone in society should look after each other and contribute to society in a meaningful way. The concept of *Folkhemmet* gave rise to the current welfare state in Sweden and it permeates the Swedish culture, especially when it comes to government regulations related to healthcare. Sweden's universal healthcare system means that the government pays a majority of healthcare costs for its inhabitants. Naturally, if the population is healthier, this would drive down healthcare costs and contribute to a sustainable healthcare system. As such, Sweden heavily prioritized predictive/preventative medicine practices in order to improve the health of society and, as a result, reduce healthcare costs. Sweden's commitment to preventive medicine is also present when it comes to the funding of scientific research. As such, there are several research groups in Sweden focused on understanding, predicting and preventing disease, some of which host students as part of this proposal. By working in these laboratories as well as by simply living in Sweden, students are exposed to the concept of *Folkhemmet*, which provide them with a unique perspective which is quite distinct from the culture in the United States.

Description of IRES Site Activities

This program supports 24 undergraduates (8/year) enrolled in PUIs across the U.S. for ten-week research experiences from June-August of 2025-27. Students travel to Stockholm to participate in cutting-edge AI and data science research projects. Additionally, students participate in extensive pre-departure training, professional development workshops, scientific and public policy events, career readiness workshops and present their work at the annual BMES meeting in October. A complete overview of the program structure and timeline is presented in Table 1.

Pre-departure training. The first step to ensuring students are able to effectively contribute to their projects begins with the assessment of the students' application materials. The PI and foreign mentors assess the candidates' academic background, personal experiences and passion for scientific research and extend offers to students who have the technical foundation and

personal maturity to succeed during their summer experience. In order to ensure that the selected IRES students have the technical skills and background knowledge necessary for success in their projects, they also participate in an 8-week pre-departure training series designed and led by the PI. This training series leverages existing experience the PI gained through the development of the pre-departure materials for a previous IRES site [10], [11], [12]. Specifically, the PI focuses on developing coding skills and knowledge of the biological systems the students will be working on. Additionally, a portion of these training sessions are dedicated to cultural training, travel logistics and preparing students to live abroad. The overall aim of this series is to prepare students to meaningfully contribute to the host lab's research and prepare them for living in Sweden. The pre-departure training series are divided into two portions, a synchronous virtual training series on Zoom and asynchronous coding tutorials, described below.

Table 1: Program Structure and Timeline

	Program Activity	Resposibility	Timing
Pre-Program Activities	Program advertising/recruitment	PI	December-January
	Application deadline	Student applicants	Februrary 1st
	Student selection and notification	PI and foreign mentors in Sweden	Februrary 28th
	Pre-program survey administration	Evaluator, selected IRES students	March
	Pre-departure training (virtual)	PI, IRES students, USD International Center	April-May
Research Experience	Welcome Weekend and Orientation	PI, IRES students	June
	Research in foreign host laboratories	PI, IRES students	June-August
	Professional Development Workshop Series	PI, IRES students	June
	Biomedical Engineering Society abstract due	IRES students	Late July
	AI and Health Open Forum	PI, foreign mentors, IRES students, invited panelists	Early August
	IRES Final Presentation Symposium	PI, foreign mentors, IRES students	August
	Final research posters due	IRES students	End of August
Post-Program Activities	Career Readiness Workshop Series (virtual)	PI, IRES students	September-October
	Biomedical Engineering Society Meeting	PI, IRES students	October
	Post-program evaluation survey administration	Evaluator, IRES students	End of October
	Program evaluation/assesment	Evaluator	November
	Review of assessment results	PI, Evaluator, foreign mentors	December

Synchronous virtual training series: The PI leads weekly 90-minute Zoom sessions for 8 weeks in April-May in the leadup to the summer research experience. These Zoom sessions consist of lectures/activities centered around: program logistics and expectations, conducting scientific research, living in Sweden, Swedish culture (inside and outside the lab), AI, data science, and algorithm bias. Four of the Zoom sessions in the training series are dedicated to student-led journal club discussions where students present a paper published by their host lab and field questions from the PI and peers. The journal club activity is designed to teach IRES students the methods, background and vocabulary that serves as the basis for their summer research project.

Asynchronous coding tutorials: All students admitted to the program have previous computer programming experience, but additional training materials is assigned to ensure student success. Given that students work with computational techniques specific to their projects, individually tailored tutorials are curated/designed by the host labs in collaboration with the PI. Developing these tutorials takes advantage of existing materials (i.e. DataCamp and course materials developed by the PI and/or host labs). Students work on these tutorials independently for two months prior to the research experience and the PI is available to address any questions.

Welcome Weekend & Orientation. After arriving in Sweden, the PI hosts a welcome weekend in Stockholm to build comradery within the cohort and introduce students to the city. The PI hosts an orientation session at KI on the Monday following arrival in Sweden.

Professional Development Workshop Series. While the PI is in Stockholm during the first 3 weeks of the program, he runs a 2-hour weekly professional development workshop series

focused on providing students with the skills necessary to succeed in their summer research project and beyond. Active-learning workshop activities include: scientific communication (delivering an elevator pitch for different audiences, presenting at lab meeting, writing a conference abstract, making and presenting a scientific poster), research methods (hypothesis generation, laboratory documentation, data management, searching the scientific literature), and identifying the broader impacts of their research (scientific outreach, research ethics and algorithm bias). In addition to developing students' research skills, a major goal of this workshop series is to develop a sociotechnical mindset in the students. The concept behind developing a sociotechnical mindset comes from the fact that engineering and computational problems are deeply embedded in the real social and political world [13]. As such, a workshop session is dedicated to an activity centered around algorithm bias. In running this workshop, the PI draws upon previous work published as part of a course taught at the University of San Diego [14].

IRES Program Blog. Throughout the summer, students write weekly blog posts on the program's website. These blog posts serve as a venue for public engagement in the students' research and overall experience abroad. Additionally, this blog helps develop students' popular scientific communication skills, an important skill to learn for scientifically engaging with the public.

Final Presentation Symposium. During the last week of the program, students give 15-minute presentations to their peers and research mentors. This event is aimed at developing students' technical presentation skills. Presentation recordings are posted on the IRES website to facilitate public engagement with the program. Students also turn in a final poster to the PI and host lab.

Career Readiness Workshop Series. After the research program, students participate in a virtual workshop series to prepare them for their future careers with a focus on encouraging and preparing them for graduate school. The PI leads weekly 90-minute Zoom sessions for 6 weeks in September-October in the leadup to the BMES conference. Workshop topics/activities include: preparing scientific manuscripts based on their summer work, resume writing, motivations for going to graduate school, graduate school applications (i.e. identifying graduate schools, graduate fellowships, personal statements, etc.), networking, presenting at and attending conferences (aimed at preparing students for the BMES conference), and a Ph.D. career panel to learn about different STEM career paths.

*Biomedical Engineering Society Meeting***.** Following the summer program, students present their work at the BMES annual meeting in October. This is an opportunity for the students to network with researchers and speak with potential graduate school mentors. Additionally, many graduate programs host receptions during the conference directed at recruiting undergraduate students to their programs. IRES students attending the conference speak with faculty members at these receptions to learn more about what each program has to offer.

Continual improvement. In order to ensure continual improvement, an external evaluation takes place each year.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award No. 2420210.

References Cited

[1] C. Manzoni *et al.*, "Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences," *Brief. Bioinform.*, vol. 19, no. 2, pp. 286–302, Mar. 2018, doi: 10.1093/bib/bbw114.

[2] L. Wen *et al.*, "Single-cell technologies: From research to application," *Innov. Camb. Mass*, vol. 3, no. 6, p. 100342, Nov. 2022, doi: 10.1016/j.xinn.2022.100342.
[3] S. Weisenburger and V. Sandoghdar, "Light microscopy: an ongoing contemporary revolution," *Contemp. Phys.*, vol. 56, no. 2, pp. 123–143, Apr. 2015, doi: 10.1080/00107514.2015.1026557.
[4] J. H. Thrall *et al.*, "Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success," *J. Am. Coll. Radiol.*, vol. 15, no. 3, pp. 504–508, Mar. 2018, doi: 10.1016/j.jacr.2017.12.026.
[5] L. Seyyed-Kalantari, G. Liu, M. McDermott, I. Y. Chen, and M. Ghassemi, "CheXclusion: Fairness gaps in deep chest X-ray classifiers," in *Biocomputing 2021*, Kohala Coast, Hawaii, USA: WORLD SCIENTIFIC, Nov. 2020, pp. 232–243. doi: 10.1142/9789811232701_0022.
[6] L. Seyyed-Kalantari, H. Zhang, M. B. A. McDermott, I. Y. Chen, and M. Ghassemi, "Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations," *Nat. Med.*, vol. 27, no. 12, pp. 2176–2182, Dec. 2021, doi: 10.1038/s41591-021-01595-0.
[7] A. J. Larrazabal, N. Nieto, V. Peterson, D. H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," *Proc. Natl. Acad. Sci.*, vol. 117, no. 23, pp. 12592–12594, Jun. 2020, doi: 10.1073/pnas.1919012117.
[8] A. Arora *et al.*, "The value of standards for health datasets in artificial intelligence-based applications," *Nat. Med.*, vol. 29, no. 11, Art. no. 11, Nov. 2023, doi: 10.1038/s41591-023-02608-w.
[9] K. Johannisson, "The People's Health: Public Health Policies in Sweden," in *The History of Public Health and the Modern State*, Brill, 1994, pp. 165–182. doi: 10.1163/9789004418363_006.
[10] M. A. Chapman, "Work in Progress: Development of a Training Program to Prepare Students for an Immersive Bioinformatics Summer Research Experience," presented at the 2021 ASEE Virtual Annual Conference Content Access, Jul. 2021. Accessed: Jan. 18, 2024.
[11] M. Chapman, "US-Sweden Bioinformatics NSF-IRES Year 1: Program Development and Initial Lessons Learned," presented at the 2022 ASEE Annual Conference & Exposition, Aug. 2022.
[12] M. A. Chapman, "Board 427: US-Sweden Bioinformatics IRES Year 2: Improving Student Preparedness," presented at the 2023 ASEE Annual Conference & Exposition, Jun. 2023. Accessed: Jan. 18, 2024.
[13] D. A. Chen, M. H. Forbes, G. D. Hoople, S. M. Lord, and J. A. Mejia, "The 'Who' in Engineering: Sociotechnical Engineering as Memorable and Relevant," *Int. J. Eng. Pedagogy IJEP*, vol. 13, no. 5, Art. no. 5, Jul. 2023, doi: 10.3991/ijep.v13i5.36571.
[14] D. A. Chen, M. A. Chapman, and J. A. Mejia, "Balancing Complex Social and Technical Aspects of Design: Exposing Engineering Students to Homelessness Issues," *Sustainability*, vol. 12, no. 15, Art. no. 15, Jan. 2020, doi: 10.3390/su12155917.