Education challenges among early career researchers in Medical Informatics

Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240701

Education Challenges Among Early Career Researchers in Medical Informatics

Narjes ROHANI[a,1], Michael GALLAGHER[b], Kobi GAL[c] and Areti MANATAKI[d]
[a] *Usher Institute, University of Edinburgh, UK*
[b] *Moray House School of Education and Sport, University of Edinburgh, UK*
[c] *School of Informatics, University of Edinburgh, UK*
[d] *School of Computer Science, University of St Andrews, UK*
ORCiD ID: Narjes Rohani https://orcid.org/0000-0003-4441-9359

Abstract. There is limited knowledge about early career researchers' challenges when studying the interdisciplinary field of Medical Informatics (MI). We conducted a qualitative content analysis through semi-structured interviews with early career researchers in MI, including individuals pursuing Master's, PhD, and postdoctoral research programmes, across two higher education institutions in the UK. We identified five challenges, including understanding biological jargon, interpreting biological data, interdisciplinary communication, understanding mathematical/statistical concepts, and programming difficulties. These insights and suggested actions to address those challenges can help to improve MI education.

Keywords. Medical informatics, Health data science, Qualitative analysis, Medical education.

1. Introduction

Since research in the medical field has become increasingly quantitative due to the increased amounts of available data, researchers have encountered challenges in utilising computational skills and managing the interdisciplinary nature of their research [1-4]. The field of Medical Informatics (MI), which is broadly defined as using computational methods to analyse biological, medical, and/or health data, has emerged to improve human health. We considered the bioinformatics and health informatics fields under the umbrella of MI, according to the aforementioned definition [1, 5, 6].

Given the urgent need to educate health professionals who are skilled in data analysis, it is crucial to improve education in MI [1, 7]. A few research studies recognise challenges in medical education [1, 5]. For example, recently, Işık et al. [1] conducted research to investigate the grand challenges in bioinformatics education. They aimed to understand why many institutions still face difficulties in bioinformatics education. They discussed seven grand challenges, categorising them into those that affect the learner, such as identifying fundamental knowledge; those that impact the educator, such as the lack of enough knowledge; and those that are critical for the long-term strengthening of bioinformatics education. Since no research has focused on the challenges of the early

[1] Corresponding Author: Narjes Rohani, Usher Institute, Old Medical School, Teviot Place, Edinburgh, UK; E-mail: Narjes.Rohani@ed.ac.uk.

career researchers' community in MI, this study specifically focuses on identifying the challenges faced by this community, including doctoral and postdoctoral researchers, as well as postgraduate research students.

Previous studies have focused on identifying the challenges faced by undergraduate students rather than early career researchers in MI and were solely based on the opinions of educators [1]. To address these gaps, and with the aim of improving MI education, we conducted a preliminary study using qualitative content analysis to discover what challenges early career researchers encounter in MI.

2. Methods

We followed the Standards for Reporting Qualitative Research (SRQR) [8] to present our research. Our study was a qualitative content analysis (designed by all authors with knowledge in qualitative analysis), involving semi-structured interviews with early career researchers including postgraduate students (both master's and PhD students doing research in the MI field) and postdoctoral researchers, in the field of MI in two UK-based higher education and research institutions. It should be noted that except for master's students, the rest of the participants did not have mandatory taught courses. Participants were selected on the basis of carrying out research in the MI field at the time when the interviews were conducted. Ethical approval for this study was granted by the School of Informatics, University of Edinburgh. Prior to the interviews, participants were invited by email and received a document describing the research goals, the interview procedure, and ethical considerations. At the beginning of the interviews, participants were given the opportunity to ask any questions and they declared their consent through signing a consent form. The interviews were conducted in individual and in-person format from Jun. 2023 to Aug. 2023. Each interview lasted between 15 and 30 minutes, which was sufficient given the focus of this study. The interviews were carried out by the lead author (NR), who has an academic background in computer science and MI. Each interview started with asking participants about their research and then participants were asked about the challenges that they face in MI. The interviews were audio-recorded, and we continued the data collection process until obtaining data saturation (no new codes emerged). We successfully recruited 10 participants (see Table 1).

The audio-recorded interviews were transformed into text using a custom Python script that employed AssemblyAI's Python SDK [9]. Afterwards, NR checked the accuracy of the transcriptions and applied the necessary corrections. The transcriptions were imported into the NVivo 14.23.2 software [10], which was utilised for coding the transcribed interviews. The coding process was carried out iteratively, and NR and MG collaborated to enhance and refine the codes throughout the process [11].

Table 1. Characteristics of the participants recruited. It should be noted that those with mixed backgrounds, declared a basic knowledge in computational concepts.

Characteristic	Description	
Gender	Female	50%
	Male	50%
Age	Ranged from 20 to 35	
Ethnicity	Diverse regions such as Asia, Europe, the Middle East, and South America	
Academic degree	Final year Master's student	10%

	Researcher in MI (recent Master's graduate)	10%
	PhD research student	60%
	Postdoctoral researcher	20%
Academic background (their undergrad or their last degree)	Computational (e.g., computer science, statistics, and math)	60%
	Biological/Medical (e.g., medicine, biology, and genomics)	20%
	Mixed (MI-related disciplines such as bioinformatics, medical informatics, and health informatics)	20%

3. Results

In this section, we present the results of the semi-structured interviews, accompanied by quoted evidence from participants. The results indicate that the most challenging aspects for the learners are **understanding biological jargon** (mentioned by 50% of participants), **interpreting biological data and results** (50% of participants), **communicating in an interdisciplinary environment** (30% of participants), **comprehending statistics and mathematical theories** (30% of participants), and **programming** (20% of participants). **Figure 1** shows the percentage of participants that discussed each challenge, as well as the most frequently mentioned words in the interviews.

The challenge of understanding biological jargon was reported by 5 out of 10 participants a total of 9 times. Among these participants, all studied a computational-related field during their undergraduate studies, which can explain how a lack of basic knowledge in biology made understanding biological terms very challenging for them. One postdoctoral participant who had a computational background stated: *"[…] I needed to read about the recent literature and biological papers, which was really hard for me to understand every single bit of those articles. (P1)".* Likewise, another participant with computational background mentioned: *"[…] it's hard for me when I start reading a paper in this field for literature review. So, I find lots of biological keywords that I'm not familiar with […]. (P2)".* The next challenge was interpreting biological data and results, which were reported primarily by participants with computational backgrounds (4 participants) as well as by one participant with a biological background. A participant with a biological background mentioned: *"Interpreting plots and results, especially the biological aspects, was more challenging. Generating a plot is relatively simple, but extracting meaningful insights from it is more complex. (P10)".* Similarly, a participant with a computational background also stated: *"When you are generating any results, it is very difficult for you to understand those results. You need to explain those results in the format of biological. It is very very difficult if you are a computer scientist, and you are working in a biological site. (P4)".*

The next challenge reported by participants was communication in interdisciplinary settings. This was highlighted by one participant with a mixed background and two participants with computational backgrounds. A postdoctoral researcher with a computational background talked about the challenge of establishing a common language with biologists: *"When I began working in an interdisciplinary environment, I needed to communicate with the biologists directly and their way of verbalising their concepts and their intentions was completely different and to find a common language was a bit tricky for me. (P1)".* Also, a participant with a mixed background discussed collaboration challenges in the context of coding: *"The most challenging thing was to collaborate with others, especially in a coding project. We needed to work with others*

and make sure that our code was understandable to everyone. (P7)". Another challenge reported, in particular by participants with biological and mixed backgrounds, was understanding statistics and mathematical theories (mentioned by 3 participants 4 times). For instance, one participant with a mixed background and experience in MI during his master's studies mentioned: *"It was much more challenging than biology because, in computational methods, you need to understand all the statistics involved which I had very little background knowledge of. (P7)".* The final challenge was related to programming, and this was mentioned by one participant with a mixed background and one with a computational background. The participant with a mixed background noted: *"[…] You need to be able to do a lot of programming, which I had very little background knowledge of (P7)".*

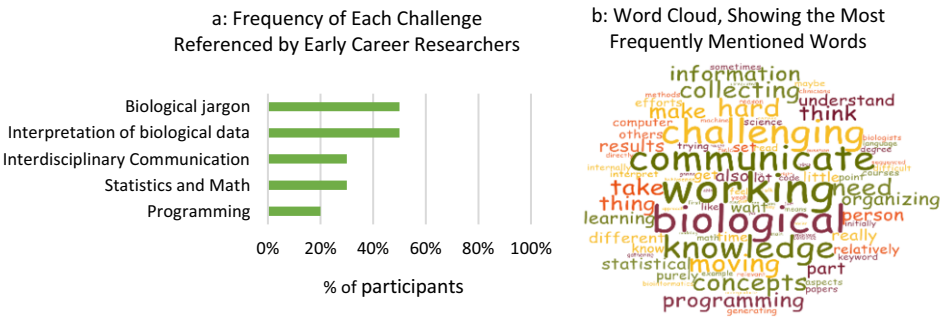

Figure 1. a: Percentage of participants discussed each challenge, b: Word cloud of the interviews, showing the most frequently mentioned words in the interviews.

4. Discussion

Learners with computational backgrounds pointed out challenges in understasnding biological jargon. Conversely, learners with biological backgrounds find statistics and mathematical concepts hard to understand, and they also find programming to be challenging. This finding is not surprising given the interdisciplinary nature of MI [1, 5]. Challenges in interpreting biological data and communicating in interdisciplinary settings were also discussed, and they are in line with existing literature [1, 12-14].

Peer learning [2, 15] in a group that includes members with diverse backgrounds, can help address difficulties related to understanding biological jargon, statistics, and foundational math knowledge. Also, peer learning can enhance communication skills in an interdisciplinary environment. To tackle programming challenges, practising and applying programming skills to solve real-world problems can be beneficial [2]. Offering foundational courses tailored to students' backgrounds and focusing on practical problem-solving may also aid in addressing these challenges.

Regarding the limitations of the study, the results are only based on self-declared data, which might be consciously or unconsciously affected by different factors, such as social desirability. To strengthen the conclusions, it would be advantageous to use quantitative methods, such as learning analytics [16]. Further studies could also expand on the small sample presented in this paper to test the generalisability of our findings.

5. Conclusions

Our results reveal that early career researchers in medical informatics face difficulties in understanding biological terms, interpreting biological data, comprehending statistics and programming, as well as communicating in interdisciplinary environments. Future work can consider the power of personalised learning to address these challenges.

Acknowledgement

This work was supported by the Medical Research Council [grant number MR/N013166/1]. We thank all the learners who participated in this study. We extend our thanks to Dr. Syed Haider and Prof. Claudio Angione for their invaluable assistance in recruiting participants.

References

[1] Işık EB, Brazas MD, Schwartz R, Gaeta B, Palagi PM, van Gelder CW, Suravajhala P, Singh H, Morgan SL, Zahroh H, Ling M. Grand challenges in bioinformatics education and training. Nature Biotechnology. 2023 Aug;41(8):1171-4.
[2] Rohani N, Gal K, Gallagher M, Manataki A. Providing Insights into Health Data Science Education through Artificial Intelligence. bioRxiv. 2024:2024-03.
[3] Sapci AH, Sapci HA. The effectiveness of hands-on health informatics skills exercises in the multidisciplinary smart home healthcare and health informatics training laboratories. Applied clinical informatics. 2017 Oct;8(04):1184-96.
[4] Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, Hovenga E, Lun KC, Marin H, Martin-Sanchez F, Wright G. Recommendations of the International Medical Informatics Association (IMIA) on education in biomedical and health informatics. Methods of information in medicine. 2010 Apr;49(02):105-20.
[5] Haux R. Medical informatics: past, present, future. International journal of medical informatics. 2010 Sep 1;79(9):599-610.
[6] Wan T, Gurupur V. Understanding the difference between healthcare informatics and healthcare data analytics in the present state of health care management. Health services research and managerial epidemiology. 2020 Aug 24;7:2333392820952668.
[7] Wee R, Soh E, Giles D. Teaching data science to medical trainees. Clinical Teacher. 2021 Aug 1;18(4).
[8] O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Academic medicine. 2014 Sep 1;89(9):1245-51.
[9] AssemblyAI. AssemblyAI's Python SDK 2023. [Accessed 2024 May 10]. Available from: https://www.assemblyai.com.
[10] Dhakal K. NVivo. Journal of the Medical Library Association: JMLA. 2022 Apr 4;110(2):270.
[11] Saldaña J. The coding manual for qualitative researchers.
[12] Tastan Bishop Ö, Adebiyi EF, Alzohairy AM, Everett D, Ghedira K, Ghouila A, Kumuthini J, Mulder NJ, Panji S, Patterton HG, H3ABioNet Consortium, as members of The H3Africa Consortium). Bioinformatics education—perspectives and challenges out of Africa. Briefings in bioinformatics. 2015 Mar 1;16(2):355-64.
[13] Schneider MV, Watson J, Attwood T, Rother K, Budd A, McDowall J, Via A, Fernandes P, Nyronen T, Blicher T, Jones P. Bioinformatics training: a review of challenges, actions and support requirements. Briefings in bioinformatics. 2010 Nov 1;11(6):544-51.
[14] Kuo MH, Sahama T, Kushniruk AW, Borycki EM, Grunwell DK. Health big data analytics: current perspectives, challenges and potential solutions. International Journal of Big Data Intelligence. 2014 Jan 1;1(1-2):114-26.
[15] Coliñir JH, Gallardo LM, Morales DG, Sanhueza CI, Yañez OJ. Characteristics and impacts of peer assisted learning in university studies in health science: A systematic review. Revista Clínica Española (English Edition). 2022 Jan 1;222(1):44-53.
[16] Rohani N, Gal K, Gallagher M, Manataki A. Early prediction of student performance in a health data science MOOC. In Proceedings of the 16th International Conference on Educational Data Mining 2023 Jul 5. International Educational Data Mining Society.