Algorithmic bias in health care: Opportunities for nurses to improve equality in the age of artificial intelligence

Siobhan O'Connor[a]*, Richard G. Booth[b]

[a]*Division of Nursing, Midwifery and Social Work, The University of Manchester, Manchester, United Kingdom*
[b]*The Arthur Labatt Family School of Nursing, Western University, London, Ontario, Canada*

Artificial Intelligence (AI) consists of a range of sophisticated computational techniques, encompassing machine learning algorithms and natural language processing among others, that are lauded as a way to improve clinical decision making, patient care, and health service delivery. A recent systematic review of AI in nursing and midwifery found many clinical, managerial, and educational applications of these predictive algorithms over the last 20 years covering areas such as wound care, critical care, falls, infection control, emergency care, older adult care, and education among others (O'Connor et al., 2022). For example, An et al. (2021) employed a number of machine learning algorithms to develop a predictive model that stratified patients admitted to intensive care units based on disease severity and care needs, while Lee et al. (2020) developed a mobile app incorporating a convolutional neural network to help predict burnout among nurses, and Narang et al. (2021) utilized a deep learning algorithm to train nurses to use an echocardiogram. Despite numerous identified benefits of AI in health care, it can also introduce a host of risks – one of the most pressing being *algorithmic bias*.

Patients of certain ethnicities, religious backgrounds, ages, and those with differing sexual preferences and genders can face discrimination when accessing health care in some countries (Dobrowolska et al., 2019; Irvin et al., 2014; Khera et al., 2014), inequalities which could be exacerbated if AI is utilized inappropriately (World Health Organization, 2020). Panch et al. (2019) defined algorithmic bias as: "*the instances when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact inequities in health systems.*" A seminal study demonstrated that a commercial risk-prediction algorithm widely used by hospitals and insurers in the United States to allocate care to thousands of patients, systematically discriminated

against Black patients (Obermeyer et al., 2019). The algorithm calculated health care needs based on total yearly health care costs drawn from electronic health records, using the assumption that this was a reasonable proxy to identify patients with complex health care needs. This led to higher risk scores being attributed to White patients who went on to receive more personalized care, as historically Black patients accessed health care less often. The study authors suggest that the mechanism of this bias may have resulted from a host of intersecting socio-political-economic factors commonly faced by Black patients, including economic barriers, historical distrust, and experiences of racial discrimination in health systems. After contacting the manufacturer about the study's findings, the research group worked with their technical teams to replicate and confirm the results, and then adjusted the prediction model to reduce the identified algorithmic bias. By partnering and collaborating with industry, researchers including nursing scientists could rigorously evaluate AI algorithms, their underlying datasets, and predictive models to help improve the development and use of AI in health care.

Algorithms deployed at scale with less oversight could have far reaching and possibly harmful effects. A systematic review by Wen et al. (2022) reviewed publicly available skin image datasets used to try and improve skin cancer diagnosis via machine learning techniques. They found the datasets largely comprised of images of lighter skin types, with darker skins underrepresented and ethnicity data underreported. The authors postulate that if predictive algorithms and models are used in cancer screening and diagnostic systems, they must be trained and tested on comprehensive datasets or at a minimum a transparent dataset. This would facilitate a critical evaluation and understanding of any AI based clinical decision support tools to avoid underperformance or misdiagnosis, and any subsequent health disparities that algorithmic bias may create due to geographic and ethnic gaps in digital health data. Moreover, Gianfrancesco et al. (2018) caution that socioeconomic data such as housing, education, and employment are often not captured in-depth in electronic health record or health insurance datasets which if used to develop AI tools could further exacerbate the problem. To reduce this risk, nurses should ensure AI tools in health care are rigorously assessed and AI research includes participants from a range of sociodemographic backgrounds where possible to address health data poverty (Ibrahim et al., 2021). Accurately reporting this data in published studies and the metadata of any open datasets generated as a result would also be important to help address algorithmic bias.

Due to nurses' close therapeutic relationships with patients and their families, they have an important role to play in educating them about AI based digital tools. However, many nurses remain unaware of the risks that AI techniques and the technologies they are based on pose such as algorithmic bias. Furthermore, nurses may not appreciate that predictive algorithms already permeate various health technology systems used locally and globally, and other popular digital platforms such as social media that many patients are exposed to. Hence, a more proactive approach is urgently needed to teach nurses about this important subject. Informatics curricula covering the fundamentals of AI computational techniques and the ethical issues it brings, such as algorithmic bias, need to be included in undergraduate and postgraduate programs to educate nursing students and the workforce (Booth, Strudwick, McBride, O'Connor, & López, 2021). This will likely have a positive impact on the development and application of AI in health care and will be essential for educating patients and their families on the advantages and disadvantages of AI-based decision support tools. The World Health Organization (WHO) (2020) recommend that those who use and are affected by AI technologies, such as patients and providers, are actively involved in their design and evaluation which nurses could support.

While many countries do not yet have extensive digital health datasets that enable AI, low- and middle-income and other nations are likely to use electronic health records and other technologies to strengthen patient care and service delivery in the future. This could generate additional complexities around algorithmic bias as the varied geo-political landscape, different interpretations of health care and nursing, and other contextual social, political, religious, and economic forces may affect the application of AI. The WHO (2020) also encourage the adoption of open-source AI software to help ensure these novel analytical techniques are developed, applied, and assessed appropriately and freely available to those in low- and middle-income countries to avoid widening the digital divide, an area that nurses working in global health can contribute to.

Some machine learning algorithms such as artificial neural networks operate as a 'black box' as the variables that influence the final predictive model are unknown. This combined with the protections afforded commercial AI tools and systems means it can be challenging to evaluate and critique how some algorithms operate and the impact they have on the decisions of professionals delivering front-line health care services as well as those in management and leadership roles. Nevertheless, nurses should promote algorithmic accountability, including lobbying local and national politicians and health care leaders for legislative and policy changes, and become more involved in the development and governance of AI in health care. Some progress is occurring, as the Algorithmic Accountability Act was introduced in the United States in 2019 to regulate the unintended consequences of automated decision systems (Manheim & Kaplan, 2019). Nurses working in policy and patient advocacy organizations should participate in efforts to regulate AI systems in health care, to find solutions that address systematic imbalances in health care data and the automation of decision support tools.

As we delve into the age of artificial intelligence, the fair and equitable treatment of patients globally should be at the forefront of the nursing profession to ensure high quality health care is provided to all. Hence, investing in nursing education which will help ensures nurses can contribute to AI research, clinical practice, and policy to help address algorithmic bias in health care.

Authors' Contributions

Siobhan O'Connor: Conceptualization, Writing − original draft preparation; *Richard G Booth*: Writing − reviewing and editing

Acknowledgments

REFERENCES

An, R., Chang, G.-M., Fan, Y.-Y., Ji, L.-L., Wang, X.-H., & Hong, S. (2021). Machine learning-based patient classification system for adult patients in intensive care units: A cross-sectional study. *Journal of Nursing Management*, doi:10.1111/jonm.13284.

Booth, R. G., Strudwick, G., McBride, S., O'Connor, S., & López, A. L. (2021). How the nursing profession should adapt for a digital future. *British Medical Journal, 373*, n1190, doi:10.1136/bmj.n1190.

Dobrowolska, B., Jędrzejkiewicz, B., Pilewska-Kozak, A., Zarzycka, D., Ślusarska, B., Deluga, A., & Palese, A. (2019). Age discrimination in health care institutions perceived by seniors and students. *Nursing Ethics, 26*(2), 443–459, doi:10.1177/0969733017718392.

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine, 178*(11), 1544–1547, doi:10.1001/jamainternmed.2018.3763.

Ibrahim, H., Liu, X., Zariffa, N., Morris, A. D., & Denniston, A. K. (2021). Health data poverty: An assailable barrier to equitable digital health care. *The Lancet. Digital health, 3*(4), e260–e265, doi:10.1016/S2589-7500(20)30317-4.

Irvin, R., Wilton, L., Scott, H., Beauchamp, G., Wang, L., Betancourt, J., & Buchbinder, S. (2014). A study of perceived racial discrimination in black men who have sex with men (MSM) and its association with health care utilization and HIV testing. *AIDS and Behavior, 18*(7), 1272–1278, doi:10.1007/s10461-014-0734-y.

Khera, R., Jain, S., Lodha, R., & Ramakrishnan, S. (2014). Gender bias in child care and child health: Global patterns. *Archives of Disease in Childhood, 99*(4), 369–374, doi:10.1136/archdischild-2013-303889.

Lee, Y.-L., Chou, W., Chien, T.-W., Chou, P.-H., Yeh, Y.-T., & Lee, H.-F. (2020). An app developed for detecting nurse burnouts using the convolutional neural networks in microsoft excel: population-based questionnaire study. *JMIR Medical Informatics, 8*(5), e16528, doi:10.2196/16528 -e16528.

Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale Journal of Law & Technology, 21*, 106–188.

Narang, A., Bae, R., Hong, H., Thomas, Y., Surette, S., Cadieu, C., & Thomas, J. D. (2021). Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiology, 6*(6), 624–632, doi:10.1001/jamacardio.2021.0185.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (American Association for the Advancement of Science), 366*(6464), 447–453, doi:10.1126/science.aax2342.

O'Connor, S., Yan, Y., Thilo, F. J., Felzmann, H., Dowding, D., & Lee, J. J. (2022). Artificial intelligence in nursing and midwifery: A systematic review. *Journal of Clinical Nursing*, doi:10.1111/jocn.16478 - *in press*.

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health, 9*(2) 010318, doi:10.7189/jogh.09.020318 -010318.

Wen, D., Khan, S. M., Ji Xu, A., Ibrahim, H., Smith, L., Caballero, J., . . . Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: A systematic review. *The Lancet (British edition), 4*(1), e64. doi:10.1016/S2589-7500(21)00252-1. Accessed October 17, 2021

World Health Organization. (2020). Ethics and governance of artificial intelligence for health. Retrieved from https://www.who.int/publications/i/item/9789240029200