Ethical and moral implications of cancer diagnosis and prognosis by machine learning models

The use of artificial intelligence in healthcare shows great potential, but also carries a number of ethical and moral implications. Particularly, the area of machine learning (ML) can have great impact in prognosis and diagnosis tasks in human medicine context, but its integration into clinical practice remains limited due to patients’ confidentiality and data privacy, peer disagreement (between model and clinician) or liability issues, to name a few [1]. The present document highlights the major ethical and moral implications of using machine learning models in cancer diagnosis and prognosis, and how they were addressed in the development of our ML model.

**Data privacy and confidentiality**

ML modelling in healthcare involves the use of data from real patients’ who have the right to privacy and confidentiality. Thus, the use of such data implies obtaining the patients’ informed consent. Moreover, personal data allowing tracing back the data to the patient should be eliminated as thoroughly as possible. In the present work, data was obtained from TCIA, a service which de-identifies and hosts a large archive of medical images of cancer accessible for public download. According to the disclaimer from the data source: “Every effort will have been made to remove private health information (PHI) from images both by the submitter and again by tested automatic de-identification processes by UAMS as required by HIPAA for data use agreements [45CFR164.514(e)(4)], i.e., appropriate safeguards to ensure that protected health information (PHI) is not used or disclosed inappropriately. Nonetheless ethical principles command all users to make no attempt to identify individuals from whatever data elements and metadata remain.” [2].

**Bias in the data**

Data may be biased towards a particular genre, geographic location, religion, race, social status, etc., which may create and/or reinforce discrimination. Data imbalances may compromise the model generalization ability and render it useless for real-life applications. Proper data collection and curation should reduce the risk of data bias. In the present work we have no way to control this kind of issue, since the dataset obtained was already de-identified.

**Model transparency and explainability**

Transparency and explainability of a model build trust in using ML in healthcare by both clinicians and patients [3]. Development of a ML model requires taking several decisions along the process, which together with the low interpretability of some models (e.g. deep learning), turn ML as a “black box” to most clinicians. Thus, any data manipulation, feature engineering, model fine tuning and evaluation criteria should be thoroughly documented to assure transparency, credibility and auditability. In the present work, a great deal of effort was put to ensure these principles of digital ethics.

**Peer disagreement and liability**

Facing a contradictory diagnostic or prognostic opinion between the model and the clinician raises several ethical concerns. Should the clinicians follow their diagnosis when they disagree with the ML outcome? How does this conflict with medical ethics? This decisions standoff requires ethical guidelines and institutional legal frameworks to ensure the patients interests are protected and also to assign responsibility in case of a wrong decision. To our knowledge these guidelines are still scarce and implemented in a case-by-case scenario [4], however, there is a pressing need to discuss and implement strategies to address liability issues if ML is to be securely used in healthcare [5]. These ethical issues, though outside the scope of this work, were considered relevant to this document.

References

[1] <https://www.cancerimagingarchive.net/disclaimer/>

[2] Michael, R. (2019). Ethical Dimensions of Using Artificial Intelligence in Health Care. AMA Journal of Ethics, 21(2), E121-124. <https://doi.org/10.1001/amajethics.2019.121>

[3] Pattarabanjird, T., McNamara, C. The clinicians’ perspectives on machine learning. *Nat Cardiovasc Res* **1**, 189–190 (2022). <https://doi.org/10.1038/s44161-022-00033-9>

[4] Alabi, R., Vartiainen, T., Elmusrati, M. (2020) Machine learning for prognosis of oral cancer : What are the ethical challenges? Proceedings of the Conference on Technology Ethics 2020 - Tethics 2020, 1-22. <https://urn.fi/URN:NBN:fi-fe2020111690467>

[5] [Texts adopted - Framework of ethical aspects of artificial intelligence, robotics and related technologies - Tuesday, 20 October 2020 (europa.eu)](https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html) (13/10/2023)

Data Citation Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., Max, D., Pais, R. C. , Qing, D. P. Y. , Roberts, R. Y., Smith, A. R., Starkey, A., Batra, P., Caligiuri, P., Farooqi, A., Gladish, G. W., Jude, C. M., Munden, R. F., Petkovska, I., Quint, L. E., Schwartz, L. H., Sundaram, B., Dodd, L. E., Fenimore, C., Gur, D., Petrick, N., Freymann, J., Kirby, J., Hughes, B., Casteele, A. V., Gupte, S., Sallam, M., Heath, M. D., Kuhn, M. H., Dharaiya, E., Burns, R., Fryd, D. S., Salganicoff, M., Anand, V., Shreter, U., Vastagh, S., Croft, B. Y., Clarke, L. P. (2015). **Data From LIDC-IDRI** [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX>