

A GOVERNMENT CALL TO CANCER DIAGNOSIS WITH A.I.

Anonymous authors

Paper under double-blind review

1 EXTENDED ABSTRACT

Cancer diagnosis with statistical modelling, or artificial intelligence (A.I), has superseded Doctor performance in 2020 (McKinney (2020)). Deployment of such models has the potential to make cancer diagnoses reach far more people while also realizing earlier discovery increasing chance of full remission. However, the lack of data sharing between public and private institutions has hindered such automatic cancer diagnoses from becoming a reality. With machine learning research already shown sufficient, governmental institutions are urged to spend resources on safely and securely gathering cancer-related datasets to develop deployment-ready cancer diagnostic models.

End-to-end development/training of deep neural networks has regularly achieved state-of-the-art results due to its ability to handle large amounts of data (Alom et al. (2019)). Majority of the cost for the development of deep learning models lies within the gathering of appropriate datasets which has resulted in institutions withholding datasets for possible capital gain. Also given fears around Doctor-patient privilege, private and public medical institutions have the incentive to retain cancer-related data. Nevertheless, research has been done on unreasonable small datasets and still shown promising results (Hu et al. (2018), Patel et al. (2020)). More data would, however, further increase accuracy scores.

The daily diagnosis of cancer patients by human Doctors already recognizes appropriate data for deep learning (medical images, patient history, etc.). Collecting and combining such data from multiple institutions would provide large enough datasets to develop very accurate cancer diagnostic models. Given the improvements in semi-supervised learning (van Engelen & Hoos (2019)) and the newly introduced quasi-open set semi-supervised learning for big data analytics (Engelbrecht & du Preez (2020)), the high cost of annotating/labelling samples is also significantly reduced. Therefore, producing cancer-related datasets simple requires the gathering of many unlabelled samples while professionally labelling only some. Doctor-patient privilege must still be taken into account when collecting and storing such patient sensitive data.

As such, we release a *government call* to gather cancer-related datasets from multiple institutions and ensuring the process remains within the privacy protection laws. This call is of utmost importance for the developing world as additional resources can not and should not be spent on the medical diagnosis industry, which has a predicted worth of 39.6 billion U.S. dollars by 2025 (Market-watch (2019)). Instead, governments are urged to collect datasets using their already available medical infrastructure. With A.I. research showing vast real-world, life-saving applications, strategies must be put in place to harness its full potential.

REFERENCES

- Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3):292, 2019.
- Emile R. Engelbrecht and Johan A. du Preez. Introduction to quasi-open set semi-supervised learning for big data analytics, 2020.
- Zilong Hu, Jinshan Tang, Ziming Wang, Kai Zhang, Ling Zhang, and Qingling Sun. Deep learning for image-based cancer detection and diagnosis- a survey. *Pattern Recognition*, 83:134–149, 2018.

Market-watch. Diagnostic or medical imaging market 2019 global industry analysis, development, revenue, future growth, business prospects and forecast to 2025: Research reports world, Oct 2019. URL <https://www.marketwatch.com/press-release/diagnostic-or-medical-imaging-market-2019-global-industry-analysis-development-rev>

Sieniek M. Godbole V. McKinney, S.M. International evaluation of an ai system for breast cancer screening, 2020.

Darshan Patel, Yash Shah, Nisarg Thakkar, Kush Shah, and Manan Shah. Implementation of artificial intelligence techniques for cancer detection. *Augmented Human Research*, 5(1):6, 2020.

Jesper E van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine Learning*, pp. 1–68, 2019.