Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220410

Applying Artificial Intelligence Privacy Technology in the Healthcare Domain

Abigail GOLDSTEEN[a,1], Ariel FARKASH[a], Micha MOFFIE[a] and Ron SHMELKIN[a]
[a] *IBM Research – Haifa, Haifa, Israel*

Abstract. Regulations set out strict restrictions on processing personal data. ML models must also adhere to these restrictions, as it may be possible to infer personal information from trained models. In this paper, we demonstrate the use of two novel AI Privacy tools in a real-world healthcare application.

Keywords. GDPR, Privacy, Machine Learning, Artificial Intelligence, Healthcare

1. Introduction

There is a known tension between the need to analyze personal data, and the need to preserve privacy of data subjects, especially in the health domain. Data protection regulations, such as GDPR, set out strict restrictions on the collection and processing of personal data. As personal information may be derived from machine learning (ML) models using inference attacks [1], models must also adhere to these requirements[2].

Many techniques have been developed recently for privacy in ML models. However, few of them have been applied in real-world settings. We demonstrate the use of two such tools in a real system for early detection of melanoma[3]: ML model anonymization and data minimization. We demonstrate their incorporation into the system data flow and architecture and show initial results on a representative medical dataset.

2. Methods

First, model-guided anonymization [2] of the training data is performed and the model is retrained on the anonymized data. Data minimization [3] is then applied to reduce the amount and granularity of newly collected data input to the model. Both tools are available in the open-source AI Privacy Toolkit[4].

The melanoma diagnostic system is presented in Figure 1. Model anonymization is applied as part of the model training phase. It does not require changes to the training procedure itself, just an extra step of anonymizing the training data and retraining the model on the anonymized data. Data minimization is applied on the final model, during or after model validation. The resulting generalizations are fed into the data management component so it can filter and generalize features before uploading to the cloud.

[1] Corresponding Author, Abigail GOLDSTEEN, IBM Research – Haifa; E-mail: abigailt@il.ibm.com.
[2] https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530
[3] https://itobos.eu/index.php
[4] https://github.com/IBM/ai-privacy-toolkit

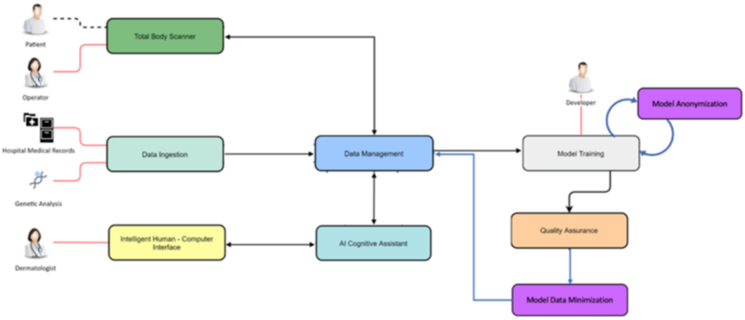

Figure 1. Melanoma diagnostic tool high-level architecture

3. Initial Results

The project has not yet begun collecting data, so results are shown on a similar dataset[5].

Table 1. Result of applying model-guided anonymization on the Hepatitis dataset with different values of k. Attack accuracy denotes the accuracy of a black-box membership inference attack on the resulting model.

Model type	k	Model accuracy	Attack accuracy
Random forest	None (original data)	0.81	0.66
	5	0.81	0.45
	10	0.78	0.46
Naïve Bayes	None (original data)	0.64	0.59
	5	0.78	0.46
	10	0.77	0.46

Table 2. Result of applying data minimization on the same dataset and models. Relative model accuracy denotes the accuracy of the model on the generalized data relative to its accuracy on the original data.

Model type	Relative model accuracy	Suppressed features
Random forest	0.98	Albumin
Naïve Bayes	0.93	Antivirals, Histology

Acknowledgements

This work has been supported by the iToBoS project funded by the European Union's Horizon 2020 research and innovation programme, under grant agreement No 965221.

References

[1] Fredrikson M, Lantz E, Jha S, Lin S, Page D, Ristenpart T. Privacy in Pharmacogenetics: An {End-to-End} Case Study of Personalized Warfarin Dosing. In23rd USENIX Security Symposium (USENIX Security 14) 2014 (pp. 17-32).

[2] Goldsteen, A., Ezov, G., Shmelkin, R., Moffie, M., Farkash, A. Anonymizing Machine Learning Models, In: Proceedings of the International Workshop on Data Privacy Management, 2021.

[3] Goldsteen A, Ezov G, Shmelkin R, Moffie M, Farkash A. Data minimization for GDPR compliance in machine learning models. AI and Ethics. 2021 Sep 26:1-5..

[5] https://archive.ics.uci.edu/ml/datasets/hepatitis