Medical Image Analysis 92 (2024) 103059

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Encrypted federated learning for secure decentralized collaboration in cancer image analysis

Daniel Truhn [a,1,*], Soroosh Tayebi Arasteh [a,1], Oliver Lester Saldanha [b,c], Gustav Müller-Franzes [a], Firas Khader [a], Philip Quirke [d], Nicholas P. West [d], Richard Gray [e], Gordon G.A. Hutchins [d], Jacqueline A. James [f,g,h], Maurice B. Loughrey [h,i,j], Manuel Salto-Tellez [f,g,h], Hermann Brenner [k,l,m], Alexander Brobeil [n,o], Tanwei Yuan [k,p], Jenny Chang-Claude [q,r], Michael Hoffmeister [k], Sebastian Foersch [s], Tianyu Han [t], Sebastian Keil [a], Maximilian Schulze-Hagen [a], Peter Isfort [a], Philipp Bruners [a], Georgios Kaissis [u,v,w], Christiane Kuhl [a], Sven Nebelung [a], Jakob Nikolas Kather [b,c,d,x]

[a] *Department of Diagnostic and Interventional Radiology, University Hospital RWTH Aachen, Aachen, Germany*
[b] *Department of Medicine III, University Hospital RWTH Aachen, Aachen, Germany*
[c] *Else Kroener Fresenius Center for Digital Health, Medical Faculty Carl Gustav Carus, Technical University Dresden, Dresden, Germany*
[d] *Division of Pathology and Data Analytics, Leeds Institute of Medical Research at St James's, University of Leeds, Leeds, United Kingdom*
[e] *Clinical Trial Service Unit, University of Oxford, Oxford, United Kingdom*
[f] *Precision Medicine Centre of Excellence, Health Sciences Building, The Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, United Kingdom*
[g] *Regional Molecular Diagnostic Service, Belfast Health and Social Care Trust, Belfast, United Kingdom*
[h] *The Patrick G Johnston Centre for Cancer Research, Queen's University Belfast, United Kingdom*
[i] *Department of Cellular Pathology, Belfast Health and Social Care Trust, Belfast, United Kingdom*
[j] *Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom*
[k] *Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany*
[l] *Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany*
[m] *German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany*
[n] *Institute of Pathology, University Hospital Heidelberg, Heidelberg, Germany*
[o] *Tissue Bank, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany*
[p] *Medical Faculty Heidelberg, Heidelberg University, Heidelberg, Germany*
[q] *Cancer Epidemiology Group, University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany*
[r] *Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany*
[s] *Institute of Pathology, University Medical Center Mainz, Mainz, Germany*
[t] *Physics of Molecular Imaging Systems, Experimental Molecular Imaging, RWTH Aachen University, Aachen, Germany*
[u] *Institute of Diagnostic and Interventional Radiology, Technical University of Munich, Munich, Germany*
[v] *Artificial Intelligence in Medicine and Healthcare, Technical University of Munich, Munich, Germany*
[w] *Department of Computing, Imperial College London, London, United Kingdom*
[x] *Medical Oncology, National Center for Tumor Diseases (NCT), University Hospital Heidelberg, Heidelberg, Germany*

ARTICLE INFO

Keywords:
Federated learning
Homomorphic encryption
Histopathology
Radiology
Artificial intelligence
Privacy-preserving deep learning

ABSTRACT

Artificial intelligence (AI) has a multitude of applications in cancer research and oncology. However, the training of AI systems is impeded by the limited availability of large datasets due to data protection requirements and other regulatory obstacles. Federated and swarm learning represent possible solutions to this problem by collaboratively training AI models while avoiding data transfer. However, in these decentralized methods, weight updates are still transferred to the aggregation server for merging the models. This leaves the possibility for a breach of data privacy, for example by model inversion or membership inference attacks by untrusted servers. Somewhat-homomorphically-encrypted federated learning (SHEFL) is a solution to this problem because only encrypted weights are transferred, and model updates are performed in the encrypted space. Here, we

* Corresponding author.
E-mail address: dtruhn@ukaachen.de (D. Truhn).
[1] Daniel Truhn and Soroosh Tayebi Arasteh contributed equally to this study.

https://doi.org/10.1016/j.media.2023.103059
Received 25 August 2022; Received in revised form 28 April 2023; Accepted 5 December 2023
Available online 7 December 2023
1361-8415/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

demonstrate the first successful implementation of SHEFL in a range of clinically relevant tasks in cancer image analysis on multicentric datasets in radiology and histopathology. We show that SHEFL enables the training of AI models which outperform locally trained models and perform on par with models which are centrally trained. In the future, SHEFL can enable multiple institutions to co-train AI models without forsaking data governance and without ever transmitting any decryptable data to untrusted servers.

One Sentence Summary:

Federated learning with somewhat homomorphic encryption enables multiple parties to securely co-train artificial intelligence models in pathology and radiology, reaching state-of-the-art performance with privacy guarantees, while requiring negligible extra computational resources.

Data availability

The data that support the findings of this study are in part publicly available, in part proprietary datasets provided under collaboration agreements. Data from the BraTS collective is publicly available under https://www.med.upenn.edu/cbica/brats2020/data.html. Data (including histological images) from the TCGA database are available at https://portal.gdc.cancer.gov/. All molecular data for patients in the TCGA cohorts are available at https://cbioportal.org. Data access for the Northern Ireland Biobank can be requested at http://www.nibiobank.org/for-researchers. All other data can be requested from the respective study groups who independently manage data access for their study cohorts.

1. Introduction

Artificial intelligence (AI) and machine learning techniques are transforming cancer imaging and cancer research and will have a profound impact on the practice of medicine(Boehm et al., 2022; Echle et al., 2021; Elemento et al., 2021; Kleppe et al., 2021). They can automate manual tasks in medical image analysis and can be used to extract hidden information from routinely available clinical image data, beyond what is visible to the human eye(Kather and Calderaro, 2020; Lu et al., 2021). AI models have been used for the detection and diagnosis of cancer, subtype classification, and optimization of cancer treatments. In particular, deep neural networks have been trained to analyze radiology images and digitized pathology slides for numerous different cancer types. For example, AI models can now detect mammographic lesions with expert-level performance(Lotter et al., 2021). Similarly, AI models predict molecular biomarkers for treatment selection directly from routine pathology slides of solid tumors(Binder et al., 2021; Coudray et al., 2018; Fu et al., 2020; Kather et al., 2020, 2019; Loeffler et al., 2022).

However, the training of AI models is infamously data hungry and requires large amounts of annotated training data. While this data may already exist, in most cases it is scattered among multiple centers. Collecting this data at a central site is hindered by obstacles which are often insurmountable in practice, most notably issues with data privacy and data governance. The data governance problem has been addressed by collaborative learning protocols such as federated learning (FL)(Lu et al., 2022; McMahan et al., 2017) in which an AI model is trained on separate sites and in which not data, but only the learned model weights are shared. This facilitates collaboration between multiple parties, but still poses significant risks for breach of patient privacy. The weight updates communicated to the central FL server contain information about the data that can be extracted to reconstruct sensitive patient information(Kaissis et al., 2021). This can be exploited through privacy attacks such as model inversion(Kaissis et al., 2020; Usynin et al., 2021; Wang et al., 2019), in which a malicious server eavesdropper captures the weight updates and attempts to recover the private dataset used to train the model or reveal other private attributes. Thus, secure multi party computation (SMPC)(Canetti et al., 2002) methods are needed by the medical community.

1.1. Prior work on privacy-preserving federated learning

One measure to protect against privacy breaches is differential privacy (DP) in which deliberate noise is added to the training updates by each site(Dwork and Roth, 2013; Kaissis et al., 2020; Truex et al., 2019). However, while this paradigm protects private information, it comes at a utility tradeoff and can lead to less performant AI models as demonstrated recently(Lu et al., 2022; Tayebi Arasteh et al., 2023). Another privacy-preserving technique which could be used for SMPC is homomorphic encryption (HE). HE can protect against a malicious server eavesdropper while maintaining AI model performance by encrypting the weight updates before sending them to the central server. One of the most common methods to implement HE in machine learning is so-called fully homomorphic encryption (FHE)(Gentry, 2009), where all the operations are done in an encrypted space. A successful implementation of FHE was first shown by Cheon et al.(Cheon et al., 2017), i.e., the CKKS algorithm (named after the authors' names: Cheon, Kim, Kim, and Song) which supports computation for almost all algebraic operations. Further works(Froelicher et al., 2021; J. X. Ma et al., 2022; Sav et al., 2021; Stripelis et al., 2021; Zhang et al., 2020) built on top of CKKS by introducing other modules such as bootstrapping or new batching mechanisms to improve the performance or to save more computation time. Although guaranteeing up to a high degree of privacy, a major downside of the CKKS-based algorithms is the high compute needed to execute(Taiello et al., 2022) which leads to very high demand of computational resource for the SMPC training process, in particular for high-dimensional data. On the other hand, none of the above works employed real-world large medical datasets to support their methods and their applicability in terms of utility and computational overhead in the medical image analysis domain is unclear. Somewhat homomorphic encryption (SHE)(Damgård et al., 2012) methods, could save computational resources while still providing privacy guarantees for certain parts of the process. One of the most successful SHE protocols is the SPDZ algorithm (named after the authors' names: Damgård, Pastro, Smart, and Zakarias)(Damgård et al., 2012), and extensions thereof(Baum et al., 2020; Damgård et al., 2013; Keller, 2020), which is based on additive secret sharing and can provide low-latency SMPC because of its very fast online phase. Keller et al. (Keller et al., 2018) showed that computational time could be drastically reduced while still preserving privacy by ignoring the zero-knowledge proof of plaintext knowledge(Bendlin et al., 2011).

We propose to use an SPDZ-based algorithm, so-called somewhat-homorphically-encrypted federated learning (SHEFL). In this setup, HE is merely employed after each local training round of participating sites. The central server performs the weight aggregation on the encrypted values and the encrypted updated weights are sent back to the clients for decryption and incorporation into their models. Importantly, since the central server does not have access to the decryption key, it cannot infer any information about which calculations have been done at individual peer locations and thus cannot extract sensitive private information. In other words, all handling of the model parameters happens in the encrypted space, making homomorphic encryption an optimal tool for

low-trust environments and handling of personal health data.

1.2. System and threat models

In this study, we examined how SHEFL can be leveraged for training of competitive AI models for cancer diagnosis and detection of cancer biomarkers in radiology and pathology images. To this end, we assumed the following threat model: A mutually trusting confederation of data owners wishes to collaboratively train a model on their joint data, but neither wants to relinquish data governance. For conducting the training, the confederation makes use of a untrusted aggregation server, which we assume to honestly participate in the protocol (i.e., faithfully conduct the aggregation procedure), but attempt to extract all available information from the weight updates sent to it by the other participants ("trusted-but-curious" threat model). We evaluated the training of AI models in three retrospective multicentric settings: 1) AI models are trained with local data only 2) AI models are trained with conventional federated learning whereby no additional measure of protection against privacy-centred attacks on the updates is undertaken and 3) AI models are trained with SHEFL in a decentralized, secure and privacy-preserving manner, whereby the individual participants encrypt their weight update before transmitting it to the server. We hypothesized that the collective and secure training of AI models reaches better accuracy than training of local models and is associated with minimal risk of privacy leakage as compared to conventional FL while keeping the cost of additional training time low due to employing HE according to the SPDZ algorithm, which is only applied immediately before weight aggregation. Furthermore, we hypothesized that dropping the zero-knowledge proof requirement(Keller et al., 2018) of the SPDZ algorithm could reduce the quadratic complexity to linear, which could substantially lower the computational time.

2. Results

2.1. SHEFL guarantees data privacy compared to conventional federated learning in the untrusted central server setting

When multiple institutions collaborate in a conventional federated learning scheme, weight updates are calculated locally and are sent to a central server to be aggregated. When unencrypted weight updates are transmitted, we demonstrate that the untrusted central server can reconstruct the training images from the weight updates in a model inversion attack. In this setting we train a neural network for the detection of malignant lesions on brain MRI examinations from the brain tumor segmentation (BraTS) dataset(Bakas et al., 2018, 2017; Menze et al., 2015). We employ a realistic setting in which data is contributed by five different institutions and in which each institution performs separate weight updates only on their data. We then perform a gradient inversion attack following the approach by Zhao et al.(Zhao et al., 2020). We demonstrate that the original training images can be reconstructed after only 120 iterations - notably, before training of the underlying neural network objective has converged, see Fig. 1. This poses a serious threat and renders the whole concept of conventional federated learning vulnerable to privacy-focused attacks. To showcase that homomorphic encryption can be used to counter these attacks and to salvage patient privacy, we repeat the training procedure, but employ homomorphic encryption in which the central server only has access to the encrypted weight updates and the key is kept private by the peers. Following the same approach - no identifiable information can be extracted from the weight updates, even after eavesdropping on the weight updates for 40,000 iterations.

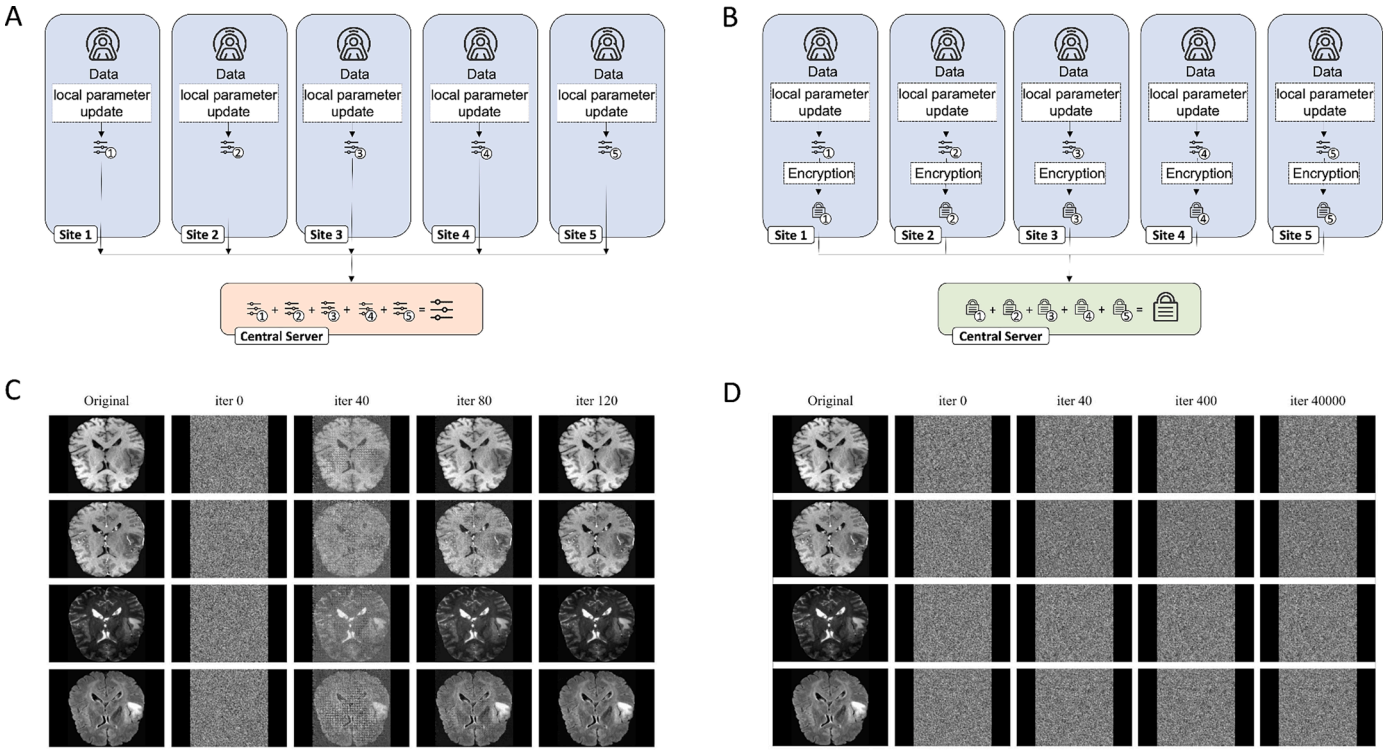

Fig. 1. Schematic of FL and SHEFL and associated Information extraction attacks. **(A)** In FL, each site trains on their own data and weight updates are transmitted to the central server for aggregation. **(B)** In SHEFL, the weight updates are encrypted and the server only has access to the encrypted values. While FL allows the server to extract patient sensitive information by reconstructing the images from the weights through gradient inversion attacks and eavesdropping on the weight updates **(C)**, this information remains protected in SHEFL and images cannot be reconstructed **(D)**. Experiments were performed on 2D slices including native T1-weighted sequences in the top row, post-contrast T1-weighted sequences in the second row, T2-weighted sequences in the third row and fluid attenuated inversion recovery sequences in the bottom row.

2.2. Secure training does not affect performance of oncological AI models

We trained AI models for tasks in oncology spanning both radiological and histopathological use-cases, see Fig. 2. Each model was trained in three settings: a) AI models are trained with local data only b) AI models are trained with conventional federated learning in a decentralized manner c) AI models are trained with SHEFL in a decentralized, secure and privacy-preserving manner. While approach a) is immune to privacy leaks, it results in training on only a limited subset of the possible data pool. Approach b) makes full use of the data but is prone to privacy leaks through the aforementioned attack by the untrusted aggregator. Only approach c) combines both training on full data and guarantees patient privacy. Moreover, as the HE scheme utilized in our study is endowed with a correctness guarantee (i.e., the values of the decrypted updates are guaranteed to be identical up to numerical precision to their plain-text counterparts), this setting does not suffer from an accuracy penalty compared to non-private training. We test the performance of each paradigm for AI models for the segmentation of glioblastoma on magnetic resonance images (MRIs) and for the detection of microsatellite instability in histopathological whole slide images (WSIs) of colorectal cancer patients.

2.2.1. Segmentation of glioblastoma on MRI

The BraTS training dataset comprises 369 MRI examinations of 369 patients which have been acquired at seventeen different clinical centers. We partitioned the data along the information where the images had been acquired into five groups and trained a 3D U-Net(Çiçek et al., 2016; Ronneberger et al., 2015) architecture to segment the tumor volume. All models were tested on an external test set from a separate institution provided by the BraTS organizers ($n = 125$) and employed the dice similarity score as a measure of performance. All five locally trained AI models performed inferior in terms of the dice score both to the models trained with FL and with SHEFL. Notably, no performance drop was seen in the model trained with SHEFL as compared to the model trained with conventional FL, cf. Table 1.

2.2.2. Prediction of genetic biomarkers in colorectal cancer patients from pathology images

In an analogous setting to the radiological use-case, we tested whether SHEFL performs equal to conventional FL and superior to locally trained models in the benchmarking task of predicting a molecular biomarker in colorectal cancer from pathology images: microsatellite instability (MSI)/mismatch repair deficiency (dMMR), which qualifies metastatic patients to receive cancer immunotherapy.

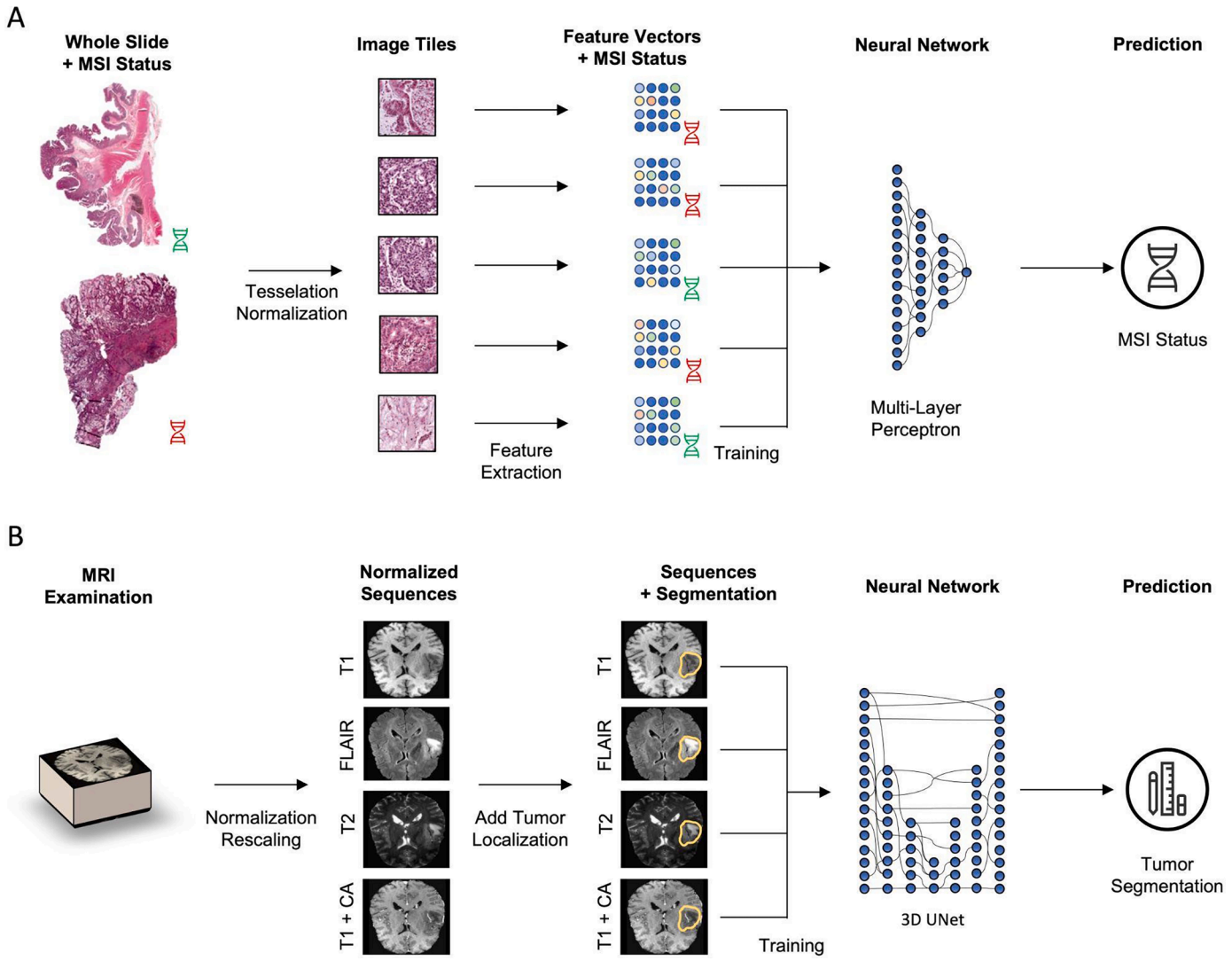

Fig. 2. Schematic of the deep learning workflow. **(A)** Histology images are first tessellated. Features are then extracted by a feature extractor network (fixed) and a multi-layer perceptron is trained to predict MSI status. **(B)** The MRI examination is normalized and rescaled to a standard resolution of 128 × 128 × 128. All four three-dimensional sequences are then fed into a 3D U-Net architecture that is trained to predict tumor segmentation outlines.

Table 1
Performance of the five radiological AI models that were trained on local data only (sites 1–5) and of the AI model that was trained with federated learning (FL) and with additional homomorphic encryption (SHEFL). P-values are given for the comparison to SHEFL.

	Site 1	Site 2	Site 3	Site 4	Site 5	FL	SHEFL
Dice Score (%)	66.15 ± 29.56 ($p < 0.001$)	78.43 ± 21.58 ($p = 0.101$)	76.58 ± 22.89 ($p = 0.021$)	77.63 ± 19.86 ($p = 0.021$)	76.54 ± 22.57 ($p = 0.003$)	81.71 ± 18.89 ($p = 0.091$)	80.32 ± 19.40

We performed the evaluation on independent test sets never seen during training: the clinical trial cohort QUASAR ($n = 1774$ patients from the United Kingdom) and the population-based cohort YCR BCIP (Yorkshire Cancer Research Bowel Cancer Improvement Programme, $n = 889$ patients). We trained three models on the Epi700 data (United Kingdom, $n = 607$), the DACHS data (Germany, $n = 2039$) and the TCGA data (USA, $n = 426$) respectively. Subsequently, we trained one model each in the federated learning setup including all three datasets without and with homomorphic encryption. Training with SHEFL was superior to training just with local data and non-inferior to training with FL both for testing on the YCR cohort and for testing on the QUASAR cohort. Both FL and SHEFL performed on the same level with no detectable difference, cf. Table 2.

2.3. Secure training is time-efficient

A notable drawback of homomorphic encryption is its computational overhead. In our study, we eschewed this drawback by encrypting not the entire training process, but only the privacy-critical weight aggregation step, which is performed by a (potentially untrusted third party), thus enabling substantial computational savings. To determine the effect of our scheme on training time compared to FL without encryption, we conducted the following experiments on a typical hardware setup used in machine learning. As a side note, de- and encryption as well as weight aggregation is usually conducted on the central processing unit (CPU), while backpropagation during training of the networks depends on the graphics processing unit (GPU).

We found that the time required for encryption was almost negligible compared to the time required to perform the backpropagation steps and the application of weight updates: for the radiological use-case described above, less than 1 % of computational time was spent on decryption, encryption and homomorphic aggregation of the weights (Fig. 3d). For the histopathological use-case, less than 5 % of time was used for decryption and encryption (which happens at edge) and homomorphic aggregation of the weights (which happens at the central server, Fig. 3b). This difference is due to the different network architectures and different number of parameters: the histopathological use-case employs a fixed backbone feature extractor(Saldanha et al., 2022) and thus has fewer parameters to optimize. Encryption and decryption scales approximately linear with the number of weights to be updated, while neural network training complexity scales more than linearly in our setup. Thus, more complex networks, such as the one used to segment brain tumors invest more computational resources in the backpropagation algorithm relative to the encryption algorithm. This is encouraging, since the relationship between training time and aggregation time is in favor of more complex networks that are usually employed when working with big data.

Table 2
Area under the receiver operating characteristic curve for the histopathological AI models that were trained for MSI detection on the Epi700, DACHS and TCGA datasets respectively and tested on the independent QUASAR and YCR-BCIP cohorts. P-values are given for the comparison to SHEFL.

	Train on Epi700	Train on DACHS	Train on TCGA	FL	SHEFL
Testing on QUASAR	74.66 ± 1.50 ($p = 0.008$)	70.38 ± 1.22 ($p < 0.001$)	70.94 ± 1.68 ($p < 0.001$)	78.52 ± 1.34 ($p = 0.289$)	79.54 ± 1.45
Testing on YCR-BCIP	77.13 ± 1.74 ($p < 0.001$)	82.46 ± 2.08 ($p = 0.054$)	78.83 ± 1.67 ($p < 0.001$)	85.42 ± 1.63 ($p = 0.270$)	86.77 ± 1.65

3. Discussion

AI has an indisputable potential in the field of oncology(Bhinder et al., 2021) and AI models are currently reaching a stage in which they can improve patient care and render medical processes more efficient (Killock, 2020; McKinney et al., 2020).

However, this improvement critically depends on the availability of sufficiently large, curated, and representative training data(Willemink et al., 2020). Currently, most research groups and industry have limited and only local data access. To train useful and generalizable AI models, stakeholders need to be able to collaborate on a large scale without jeopardizing patient privacy(Bhinder et al., 2021). Only through such multi-institutional collaboration can robust AI models be trained that adequately capture the entire human population and that make the transition from bench to bedside(Bhinder et al., 2021). Federated learning was initially proposed as a technical solution for privacy-preserving distributed AI(Konečný et al., 2017). FL enables joint training of AI models by multiple partners who do not share their data with each other and has been demonstrated to facilitate the training of AI models on big data(Dayan et al., 2021). Similarly, swarm learning (SL) utilizes a network of nodes to jointly train a model on distributed data and to aggregate model weights without a central instance(Saldanha et al., 2022; Warnat-Herresthal et al., 2021). However, FL and SL have an important shortcoming: during training, weight updates must be shared and information about the underlying data can be extracted from these weight updates as shown in our study. Such techniques should thus not be considered privacy techniques, but techniques for preserving data governance(Ziller et al., 2022). Since medical data is highly sensitive and since data privacy laws forbid the use of data in such environments, where private data can be extracted, this critically limits the applicability of collaborative learning schemes and prevents the development of powerful AI models in cancer diagnosis and treatment.

This shortcoming can be remedied by employing techniques which guarantee privacy to data owners. The only technique to guarantee privacy in a data release process is differential privacy(Dwork and Roth, 2013). Hence, when sharing the model with untrusted third parties, such a technique would have to be employed to constrain the success of attacks against patient privacy. We operate under a slightly different threat model. As all participants of the federated learning workflow described above are mutually trusting, are not intending to publish the model to the outside world and all receive an identical copy of the final model, we need only protect against an attack by the (untrusted aggregation server). Our homomorphic encryption scheme protects the weights during this critical aggregation step: local sites encrypt their weight updates before sending them out and keep the decryption key private. The entity which receives the weight updates from all sites and which performs the weight aggregation in the encrypted space thus has no access to the underlying data and no sensitive data can be extracted by design. Our technique has two notable benefits: it sidesteps the computational overhead of having to train the entire model in the encrypted space using HE. In principle, it would also be possible to use HE on all levels of the training process - i.e., also during

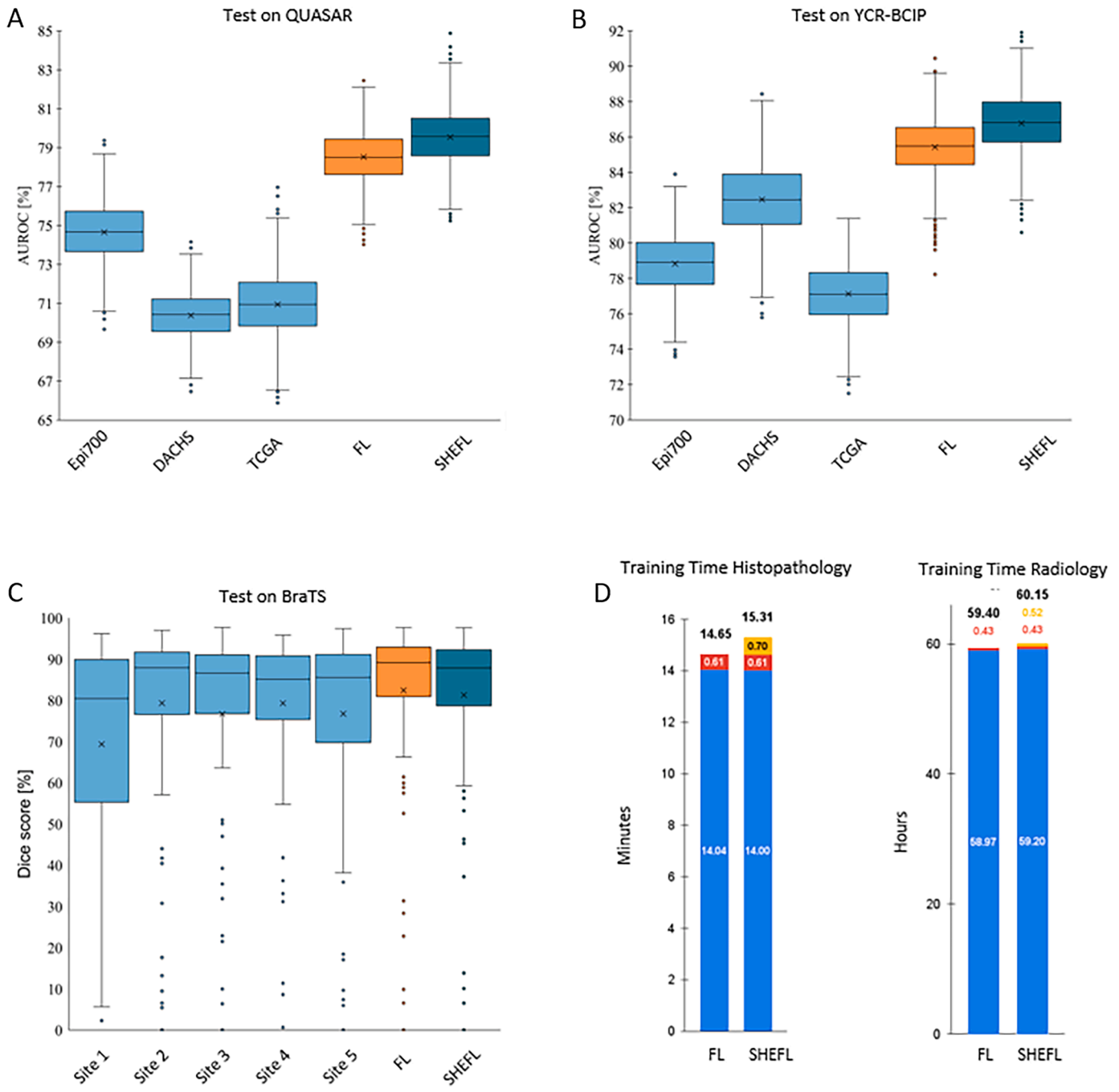

Fig. 3. Results of training on local data only vs. training using FL and SHEFL. Training neural networks on single-site datasets results in inferior performance as compared to FL and SHEFL. A neural network was trained to detect MSI on data from the Epi700, the DACHS and the TCGA cohorts respectively as well as on all three datasets using FL and SHEFL. The resulting networks were then tested on the QUASAR **(A)** and the YCR-BCIP **(B)** cohorts demonstrating superior performance of FL and SHEFL. Similarly, tumor segmentation in MRI data was trained on data from five different sites as well as on all data using FL and SHEFL. The resulting neural networks were then tested on an independent held-out test set and demonstrated improved performance **(C)**. Computational overhead was almost negligible (red: overhead for FL, yellow: additional overhead for encryption) as compared to training time needed for backpropagation (blue) **(D)**. .

backpropagation. However, with concurrently available computational resources, this has proven to be prohibitively computationally expensive and is not yet in reach(Keller et al., 2018). Furthermore, as long as all data stays on site - as is the case in our FL setup - there is no need to encrypt the backpropagation procedure: potential eavesdroppers do not have access to that part of the training procedure as it is done behind secure firewalls. By restricting the fully homomorphic encryption to the most critical part of FL - the weight aggregation - we show that additional computational overload is almost negligible. Moreover, our technique allows us to avoid the privacy-utility trade-offs of employing Differential Privacy for training, in which training with Differential Privacy can lead to less-performant AI models(Lu et al., 2022). We note that the utilization of Differential Privacy would be mandatory in threat models different from ours, especially if the final model was designed to be shared with untrusted third parties.

A similar scheme to ours was demonstrated by Kaissis et al. in a proof-of-concept study for classifying pneumonia on chest radiographs by using secure multi-party computation through additive secret sharing (Kaissis et al., 2021; Keller et al., 2018). With our study, we are the first to comprehensively assess fully homomorphic encryption in cancer

diagnosis on large multi-centric databases spanning both radiology and histopathology.

Our study demonstrates that AI models for oncological image-processing can be trained securely on multi-institutional data without compromising patient privacy. This will facilitate collaboration between researchers and industry alike, ultimately leading to the development of advanced and clinically useful AI models. We show that implementing the FL scheme together with homomorphic encryption comes with minimal additional code complexity and can be performed with our publicly available code.

A technical limitation of our study is that we performed all experiments within one institutional network. However, by distributing the datasets to different computing entities and keeping them strictly separate, we simulated the setting in which multiple institutions - each with their own network - perform FL realistically. We assumed a constant network communication cost in our experiments. In realistic settings, communication overhead can be unpredictable, as it depends on more factors than network size (such as concurrent traffic or physical distance of the sites). We thus chose to exclude this factor, believing it to only represent a minor limitation. We note that homomorphically encrypted weights cannot be efficiently reduced in size by compression, however this limitation is negligible compared to the requirement to encode them as 64-bit data types for transmission over the hypertext transfer protocol (HTTP). Moreover, as all parties are mutually trusting and receive an identical copy of the fully trained model at the end of training, we utilized the same key pair to encrypt the weights on all participating nodes, thus avoiding the technical challenge of key distribution.

Further improvements to the FL process are possible: with increasing peer numbers who participate in the FL setup, participation of a bounded number of malicious participants who try to corrupt the training process by delivering adversarial weight updates is possible, whereas we regarded all participants as either fully trusted or honest but curious. It has been shown that regular FL fails to converge in the presence of faulty and malicious clients(Blanchard et al., 2017). Measures to counter these attacks are available and can be integrated in federated learning schemes should the need arise(X. J. Ma et al., 2022).

In conclusion, our study provides a blueprint for the secure and privacy-preserving multi-institutional training of oncological AI models and solves an urgent need, since it is becoming increasingly clear that differences in race and gender affect disease risk among individuals and that existing datasets at local institutions are insufficient to account for these effects.

4. Methods

4.1. Ethics statement

This study was carried out in accordance with the Declaration of Helsinki. This study is a retrospective analysis of publicly available anonymized MRI examinations and of anonymized histopathological tissue samples from multiple cohorts of cancer patients. Collection and anonymization of patients in all cohorts took place in each contributing center. Approval by the local ethics committee at each contributing center was given if applicable (QUASAR: North East – York Research Ethics Committee; YCR: Ethical approval was not required, because screening was recommended in all patients diagnosed with CRC. Testing was considered part of the 'standard of care' clinical pathway; Epi700: Northern Ireland Biobank (NIB13/0069, NIB13/0087, NIB13/0088 and NIB15/0168), DACHS: Ethics committee of the Medical Faculty, University of Heidelberg). Approval of the ethics committee at the University Hospital of Aachen was given for the retrospective analysis of anonymized image data under reference number "Ethikkommission EK 028/19".

4.2. Patient cohorts

MRI data for the BraTS patient collective contains brain MRI scans of 341 patients collected from 17 imaging centers and additional 28 patients for whom the imaging centers were not specified by the data provider. During federated learning we allocated the patients to five data clusters simulating the situation in which a regional hospital's image database contains MRI data of multiple imaging centers. This situation is typical in real-world scenarios where patients are referred for surgery and bring their image data that had been acquired at an external institution before. The allocation of patients is detailed in supplemental Table S1. All MRI examinations contained pre- and post-contrast T1-weighted sequences, T2-weighted sequences and fluid attenuation inversion-recovery sequences (FLAIR). All sequences were acquired in axial orientation. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists.

For the histopathological data we collected digital whole slide images (WSI) of H&E-stained slides of human colorectal cancer (CRC) from five patient cohorts, three of which were used as training cohorts and two of which were used as test cohorts following the division of data in a previous study(Saldanha et al., 2022). The training cohorts are representative of real-world clinical settings. First, the Northern Ireland Epi700 ($n = 661$) cohort study contained data of patients with stage II and III colon cancer. This data was provided by the Northern Ireland Biobank(Lewis et al., 2018; Loughrey et al., 2021) (application NIB20–0346). Second, the "Darmkrebs: Chancen der Verhütung durch Screening" study (DACHS, $n = 2448$) is a large population-based case-control study. This data includes samples of CRC patients at any disease stage. This data was collected from over 20 hospitals in Germany. Data collection was coordinated by the German Cancer Research Center (DKFZ, Heidelberg, Germany)(Brenner et al., 2006; Carr et al., 2020; Li et al., 2022) and supported by the NCT tissue bank at the National Center for Tumor Diseases and the Institute of Pathology at the University of Heidelberg. Third, "The Cancer Genome Atlas" (TCGA) CRC cohort ($n = 632$) is a large collection of tissue specimens from multiple populations across different countries, but largely from the United States of America (USA) ("GDC," n.d.).

We employed two separate test cohorts: The "Quick and Simple and Reliable" (QUASAR) cohort was derived from a clinical trial of adjuvant therapy containing 2206 WSI, which aimed to determine survival benefit from adjuvant chemotherapy in CRC patients from the United Kingdom (UK)(Hutchins et al., 2011; Quasar Collaborative Group et al., 2007). The second test cohort used data from the Yorkshire Cancer Research Bowel Cancer Improvement Programme(Taylor et al., 2019) (YCR-BCIP) cohort ($n = 889$). This was a population-based study collected in the Yorkshire Region in the UK. For all cohorts, microsatellite instability (MSI) / mismatch repair deficiency (dMMR)(Marks and West, 2020) data were acquired.

The distribution of tumor stages in TCGA, DACHS and YCR-BCIP is comparable, see supplemental Table S2. In QUASAR, stage III tumors are overrepresented due to the fact that adjuvant therapy is mainly performed in intermediate stage tumors. Therefore, following previous work(Saldanha et al., 2022), we used YCR-BCIP and QUASAR as test cohorts to investigate the robustness of the AI models both on a general population and on a clinical trial population. Importantly, neither in the MRI data nor in the histopathological data, there was any overlap between training and test cohorts.

4.3. Deep learning training procedure

4.3.1. Hardware

The hardware used in our experiments were Intel CPUs with 18 cores and 32 GB RAM and Nvidia RTX 6000 GPUs with 24 GB memory.

4.3.2. MRI data

All of the 3D volumes were cropped around the brain to lower the computational costs and standardize the field of view. As intensity distributions vary across magnetic resonance images, intensity normalization is crucial. Therefore, we clipped the intensity values above the 99 percentiles of the image, then subtracted the minimum value of the result from voxel values and divided the shifted image by the maximum value of the image. We performed data augmentation during training by applying random cropping of patches of $128 \times 128 \times 128$ from each original volume around its center. Additionally, we applied medio-lateral and cranio-caudal flipping with a probability of 0.4. Intensity was randomly rescaled according to a power-law $I_{new} = g.I^{\gamma}$ (Cirillo et al., 2021) with gain g and the exponent γ randomly selected between 0.8 - 1.2 from a uniform distribution. White Gaussian noise with zero mean and a standard deviation of 0.03 was added to each sequence of the multi modal MRI data.

A modified 4-level 3D U-Net(Çiçek et al., 2016; Ronneberger et al., 2015) was utilized for segmentation of brain tumors. In the contraction path, each layer contained two $3 \times 3 \times 3$ convolutions, each followed by a rectified linear unit (ReLU)(Agarap, 2019), a batch normalization (BN) (Ioffe and Szegedy, 2015) and then a $2 \times 2 \times 2$ max pooling with strides of two in each dimension. The output channel number was doubled after each level in the contraction path, and it was 48 at the end of level one. In the expansion path, each layer consisted of a nearest neighbor up-sampling of $2 \times 2 \times 2$ in each dimension, followed by two $3 \times 3 \times 3$ convolutions each followed by a ReLU and BN. The output channel number was halved after each level in the expansion path. In the last layer, a $1 \times 1 \times 1$ convolution, which reduced the number of output channels to 3, followed by a SoftMax layer, was used for the per-voxel final classification.

The model was optimized using the Adam optimizer(Kingma and Ba, 2017) with a learning rate of 10^{-4}. To be consistent in our comparison scenarios, all the weight and bias parameters of all the different models were initialized using the He initialization scheme(He et al., 2015). As a loss function, we chose the Dice loss tailored to the BraTS data needs (Henry et al., 2021). To minimize the overhead and make maximum use of the graphics processing unit memory, we utilized large input tiles over a large batch size and reduced the batch to a single 3D image(Ronneberger et al., 2015) with 4 channels, each channel being one of the MR modalities. Hence, the batch normalization acted like instance normalization in our implementation. The network contained a total of 5,670, 579 trainable parameters.

4.3.3. Histopathological data

For prediction of molecular features from image data, we based our analysis on a well-established weakly-supervised end-to-end prediction pipeline, which was described and evaluated in a recent benchmark study(Ghaffari Laleh et al., 2022). As a preprocessing step, the original gigapixel WSIs were tessellated into patches of size $(512 \times 512 \times 3)$ pixels and were color-normalized with the Macenko method(Macenko et al., 2009). Blurry patches and patches with no tissue were removed from the data set using canny edge detection(Ghaffari Laleh et al., 2022). Following that approach, we obtained a normalized edge image using the "canny" method in Python's OpenCV(Culjak et al., 2012) package and then removed all tiles with a mean value below a threshold of 4. A pre-trained ResNet18 was used to extract a (512×1) feature vector from 150 randomly selected patches for each patient [9]. Before training, the number of tiles in each class were equalized by random undersampling until all classes had the same number of tiles, as described before(Kather et al., 2020, 2019). Feature vectors served as input to a fully connected classification network and the patient-wise MSI label was used to label every single tile derived from that patient. The fully connected classifier network comprised four layers with (512×256), (256×256), (256×128) and (128×2) connections with a ReLU activation function and the network contained a total of 492,930 trainable parameters. The model was optimized using the Adam optimizer(Kingma and Ba, 2017) with a learning rate of 4×10^{-5} and the He initialization scheme(He et al., 2015) was employed. Cross-entropy was chosen as the loss function and the model was trained in batches of size 124 for 100 epochs and utilizing 5-fold cross-validation.

4.4. Somewhat-homomorphically-encrypted federated learning (SHEFL) process

4.4.1. The collaborative learning procedure

Every participating site performed a complete local training round, in a conventional non-privacy-preserving machine learning manner, using their own data, where in our case each round equaled an epoch, leading to calculation of local gradient updates of the network parameters. Afterward, the local sites applied a homomorphic encryption setup using a public key on their gradient updates according to the SPDZ algorithm(Damgård et al., 2012) while ignoring the zero-knowledge proof of plaintext knowledge(Bendlin et al., 2011) requirement. The encrypted network parameters were aggregated according to the FedAvg (McMahan et al., 2017) algorithm by the central server in the encrypted space, leading to one set of global network parameters (which are still in the encrypted space). A copy of the global encrypted parameters was transferred back to the local sites by the central server. Using the public key, each site decrypted the global model and started another local training round with these new model parameters. This iterative process continued until the convergence of the global model.

4.4.2. Details of the homomorphic encryption method: the SPDZ algorithm

The algorithm utlizes an additive secret sharing strategy, where a message x is encrypted through distributing it as different shares to the participants. Assuming trusted-but-curious aggregation server, it requires only one crypto provider for dividing the shares between local sites. Particularly, assuming there are n sites, where $n \in \{1, 2, 3, \ldots, N\}$, each site gets assigned a random integer number in the range of $(0, Q)$ as its secret share x_n, except for the site N which gets a share as follows:

$$x_N = (x - x_1 - x_2 - \ \ldots \ - x_{N-1}) \ \% \ Q \tag{1}$$

The public key Q is a large prime number generated by the crypto provider. Consequently, the secret x could be decrypted according to Eq. (2):

$$x = (x_1 + x_2 + \ \ldots \ + x_N) \ \% \ Q \tag{2}$$

Although all the sites have access to the public key Q, none of them would know about the actual secret x as it is shared additively among them. Importantly, since the central server does not have access to Q, it cannot infer any information about the secret x. Moreover, the scheme has a homomorphic property. Thus, a certain numer of operations could be performed in the encrypted space without any information loss such as addition and multiplication. This method particularly suits our goals as we intended to solely use the HE during the weight aggregation which eventually requires only two types of operations namely addition and multiplication, i.e., no need for expensive operations such as convolution, pooling, and derivation.

Of note, this additive secret sharing algorithm assumes all numbers to be of integer values, which is in conflict with the neural network weights and biases that are usually of floating-point nature. Consequently, an important step before the encryption process is encoding the secret x into an integer value, namely using the fixed-point arithmetic (Catrina and de Hoogh, 2010; Costache et al., 2017). Subsequently, a conversion from fixed-point to the original floating-point precision happens before the decryption process. Depending on the chosen precision, this conversion could be both a lossy or a lossless process. For instance, the fractional value of 2.9874 will be represented by 2987 in the case of selecting a precision of 3. In our implementation, we

observed that a precision > 13 results in almost lossless computations for cancer image analysis when using 32-bit memory for storing the image values.

4.5. Evaluation metrics and statistical analysis

4.5.1. MRI data

The dice similarity score was employed as a measure of segmentation performance for MRI data. Statistical spread were determined for 125 points. All the mean values were accompanied by a standard deviation values. For determining statistical significance, two-tailed paired *t*-test or Wilcoxon singed-rank test were employed accounting for normality, which was tested using Shapiro-Wilk test(Shapiro and Wilk, 1965). A P-value ≤ 0.05 was considered significant.

4.5.2. Histopathological data

Area under the receiver operating characteristic curve (AUROC) was employed as the main classification evaluation metric. Bootstrapping was utilized with 1000 redraws for each measure to determine the statistical significance and spread(Konietschke and Pauly, 2014). All the mean values were accompanied by a standard deviation. A P-value $\leq$ 0.05 was considered significant.

4.6. Code availability

Our source code for secure federated learning using homomorphic encryption is publicly available at https://github.com/tayebiarasteh/federated_HE. All source codes for training and evaluation of the deep neural networks, MR image analysis and preprocessing, 3D data augmentation, and gradient inversion attack are available at https://github.com/tayebiarasteh/federated_HE. All source code for the histological image analysis is available at https://github.com/KatherLab/HIA and all source code for histological image preprocessing is available at https://github.com/KatherLab/preProcessing. All code for the experiments was developed in Python v3.8 using the PyTorch v1.4 framework. The secure federated learning process including homomorphic encryption was developed using PySyft(Ziller et al., 2021) v0.2.9.

Author contributions

JNK, DT, and STA conceptualized the study and performed the formal analysis; STA and DT developed the SHEFL software; OLS and JNK developed the histopathology image analysis software; STA and DT developed the radiology image analysis software; STA performed the methodology; STA, DT, and JNK contributed to the validation; STA and DT contributed to the visualization; JNK and DT administrated the project; PQ, NPW, RG, GGH, JAA, MBL, MST, HB, AB, TY, JCC, and MH provided histopathology resources; STA, DT, and JNK wrote the manuscript; All authors reviewed & edited the manuscript and collectively made the decision to submit for publication.

Declaration of Competing Interest

The Authors declare no competing financial or non-financial interests. For transparency, we provide the following information: JNK declares consulting services for Owkin, France, DoMore Diagnostics, Norway, Panakeia, UK, Scailyte, Switzerland, Cancilico, Germany, Mindpeak, Germany, and Histofy, UK; furthermore he holds shares in StratifAI GmbH, Germany, and has received honoraria for lectures by AstraZeneca, Bayer, Eisai, MSD, BMS, Roche, Pfizer and Fresenius. DT holds shares in StraifAI GmbH, Germany and received honoraria for lectures by Bayer. PQ and NW declare research funding from Roche and PQ consulting and speaker services for Roche. MST has recently received honoraria for advisory work in relation to the following companies: Incyte, MindPeak, MSD, BMS and Sonrai; these are all unrelated to this work. No other potential conflicts of interest are reported by any of the authors. The authors received advice from NVIDIA when performing this study, but NVIDIA did not have any role in study design, conducting the experiments, interpretation of the results or decision to submit for publication.

Data availability

The data that support the findings of this study are in part publicly available, in part proprietary datasets provided under collaboration agreements. Data from the BraTS collective is publicly available under https://www.med.upenn.edu/cbica/brats2020/data.html. Data (including histological images) from the TCGA database are available at https://portal.gdc.cancer.gov/. All molecular data for patients in the TCGA cohorts are available at https://cbioportal.org. Data access for the Northern Ireland Biobank can be requested at http://www.nibiobank.org/for-researchers. All other data can be requested from the respective study groups who independently manage data access for their study cohorts.

Acknowledgements

The authors are grateful for the support by NVIDIA who provided counsel and supported our group with two RTX6000 GPUs. We additionally acknowledge support by the tissue bank of the National Center for Tumor Diseases (NCT) at the Institute of Pathology at University Hospital Heidelberg, Heidelberg, Germany, for providing access to the biobank data.

Funding sources

DT is supported by the German Federal Ministry of Education and Research (SWAG, 01KD2215A; TRANSFORM LIVER), the European Union's Horizon Europe and innovation programme (ODELIA, 101057091). STA is funded and partially supported by the RACOON network under BMBF grant number 01KX2021. JNK is supported by the German Federal Ministry of Health (DEEP LIVER, ZMVI1-2520DAT111) and the Max-Eder-Programme of the German Cancer Aid (grant #70113864). The DACHS study (HB, JC-C and MH) was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HO 5117/2-2, HE 5998/2-1, HE 5998/2-2, KL 2354/3-1, KL 2354/3-2, RO 2270/8-1, RO 2270/8-2, BR 1704/17-1 and BR 1704/17-2), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT; Germany) and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). The Epi700 creation was enabled by funding from Cancer Research UK (C37703/A15333 and C50104/A17592) and a Northern Ireland HSC R&D Doctoral Research Fellowship (EAT/4905/13). PQ and NPW are supported by Yorkshire Cancer Research Programme grants L386 (QUASAR series) and L394 (YCR BCIP series). PQ is a National Institute of Health Research senior investigator. JAJ has received funds from Health and Social Care Research and Development (HSC R&D) Division of the Public Health Agency in Northern Ireland (R4528CNR and R4732CNR) and the Friends of the Cancer Centre (R2641CNR) for development of the Northern Ireland Biobank.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2023.103059.

References

Agarap, A.F., 2019. Deep learning using rectified linear units (ReLU).

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C., 2017. Advancing The cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data 4, 170117. https://doi.org/10.1038/sdata.2017.117.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., Prastawa, M., Alberts, E., Lipkova, J., Freymann, J., Kirby, J., Bilello, M., Fathallah-Shaykh, H., Wiest, R., Kirschke, J., Wiestler, B., Colen, R., Kotrotsou, A., Lamontagne, P., Marcus, D., Milchenko, M., Nazeri, A., Weber, M.A., Mahajan, A., Baid, U., Gerstner, E., Kwon, D., Acharya, G., Agarwal, M., Alam, M., Albiol, Alberto, Albiol, Antonio, Albiol, F.J., Alex, V., Allinson, N., Amorim, P.H., Amrutkar, A., Anand, G., Andermatt, S., Arbel, T., Arbelaez, P., Avery, A., Azmat, M., Pranjal, B., Bai, W., Banerjee, S., Barth, B., Batchelder, T., Batmanghelich, K., Battistella, E., Beers, A., Belyaev, M., Bendszus, M., Benson, E., Bernal, J., Bharath, H.N., Biros, G., Bisdas, S., Brown, J., Cabezas, M., Cao, S., Cardoso, J.M., Carver, E.N., Casamitjana, A., Castillo, L.S., Catà, M., Cattin, P., Cerigues, A., Chagas, V.S., Chandra, S., Chang, Y.J., Chang, S., Chang, K., Chazalon, J., Chen, S., Chen, W., Chen, J.W., Chen, Z., Cheng, K., Choudhury, A.R., Chylla, R., Clérigues, A., Colleman, S., Colmeiro, R.G.R., Combalia, M., Costa, A., Cui, X., Dai, Z., Dai, L., Daza, L.A., Deutsch, E., Ding, C., Dong, C., Dong, S., Dudzik, W., Eaton-Rosen, Z., Egan, G., Escudero, G., Estienne, T., Everson, R., Fabrizio, J., Fan, Y., Fang, L., Feng, X., Ferrante, E., Fidon, L., Fischer, M., French, A.P., Fridman, N., Fu, H., Fuentes, D., Gao, Y., Gates, E., Gering, D., Gholami, A., Gierke, W., Glocker, B., Gong, M., González-Villá, S., Grosges, T., Guan, Y., Guo, S., Gupta, S., Han, W.S., Han, I.S., Harmuth, K., He, H., Hernández-Sabaté, A., Herrmann, E., Himthani, N., Hsu, W., Hsu, C., Hu, Xiaojun, Hu, Xiaobin, Hu, Yan, Hu, Yifan, Hua, R., Huang, T.Y., Huang, W., Huffel, S.V., Huo, Q., Vivek, H., Iftekharuddin, K.M., Isensee, F., Islam, M., Jackson, A.S., Jambawalikar, S.R., Jesson, A., Jian, W., Jin, P., Jose, V.J.M., Jungo, A., Kainz, B., Kamnitsas, K., Kao, P.Y., Karnawat, A., Kellermeier, T., Kermi, A., Keutzer, K., Khadir, M.T., Khened, M., Kickingereder, P., Kim, G., King, N., Knapp, H., Knecht, U., Kohli, L., Kong, D., Kong, X., Koppers, S., Kori, A., Krishnamurthi, G., Krivov, E., Kumar, P., Kushibar, K., Lachinov, D., Lambrou, T., Lee, J., Lee, C., Lee, Y., Lee, M., Lefkovits, S., Lefkovits, L., Levitt, J., Li, T., Li, Hongwei, Li, W., Li, Hongyang, Li, Xiaochuan, Li, Y., Li, Heng, Li, Zhenye, Li, Xiaoyu, Li, Zeju, Li, XiaoGang, Li, W., Lin, Z.S., Lin, F., Lio, P., Liu, C., Liu, B., Liu, X., Liu, M., Liu, J., Liu, L., Llado, X., Lopez, M.M., Lorenzo, P.R., Lu, Z., Luo, L., Luo, Z., Ma, J., Ma, K., Mackie, T., Madabushi, A., Mahmoudi, I., Maier-Hein, K.H., Maji, P., Mammen, C., Mang, A., Manjunath, B., Marcinkiewicz, M., McDonagh, S., McKenna, S., McKinley, R., Mehl, M., Mehta, S., Mehta, R., Meier, R., Meinel, C., Merhof, D., Meyer, C., Miller, R., Mitra, S., Moiyadi, A., Molina-Garcia, D., Monteiro, M.A., Mrukwa, G., Myronenko, A., Nalepa, J., Ngo, T., Nie, D., Ning, H., Niu, C., Nuechterlein, N.K., Oermann, E., Oliveira, A., Oliveira, D.D., Oliver, A., Osman, A.F., Ou, Y.N., Ourselin, S., Paragios, N., Park, M.S., Paschke, B., Pauloski, J.G., Pawar, K., Pawlowski, N., Pei, L., Peng, S., Pereira, S.M., Perez-Beteta, J., Perez-Garcia, V.M., Pezold, S., Pham, B., Phophalia, A., Piella, G., Pillai, G., Piraud, M., Pisov, M., Popli, A., Pound, M.P., Pourreza, R., Prasanna, P., Prkovska, V., Pridmore, T.P., Puch, S., Puybareau, É., Qian, B., Qiao, X., Rajchl, M., Rane, S., Rebsamen, M., Ren, H., Ren, X., Revanuru, K., Rezaei, M., Rippel, O., Rivera, L.C., Robert, C., Rosen, B., Rueckert, D., Safwan, M., Salem, M., Salvi, J., Sanchez, I., Sánchez, I., Santos, H.M., Sartor, E., Schellingerhout, D., Scheufele, K., Scott, M.R., Scussel, A.A., Sedlar, S., Serrano-Rubio, J.P., Shah, N.J., Shah, N., Shaikh, M., Shankar, B.U., Shboul, Z., Shen, Haipeng, Shen, D., Shen, L., Shen, Haocheng, Shenoy, V., Shi, F., Shin, H.E., Shu, H., Sima, D., Sinclair, M., Smedby, O., Snyder, J.M., Soltaninejad, M., Song, G., Soni, M., Stawiaski, J., Subramanian, S., Sun, L., Sun, R., Sun, J., Sun, K., Sun, Y., Sun, G., Sun, S., Suter, Y.R., Szilagyi, L., Talbar, S., Tao, D., Tao, D., Teng, Z., Thakur, S., Thakur, M.H., Tharakan, S., Tiwari, P., Tochon, G., Tran, T., Tsai, Y.M., Tseng, K.L., Tuan, T. A., Turlapov, V., Tustison, N., Vakalopoulou, M., Valverde, S., Vanguri, R., Vasiliev, E., Ventura, J., Vera, L., Vercauteren, T., Verrastro, C., Vidyaratne, L., Vilaplana, V., Vivekanandan, A., Wang, G., Wang, Q., Wang, C.J., Wang, W., Wang, D., Wang, R., Wang, Y., Wang, C., Wang, G., Wen, N., Wen, X., Weninger, L., Wick, W., Wu, S., Wu, Q., Wu, Y., Xia, Y., Xu, Y., Xu, X., Xu, P., Yang, T.L., Yang, X., Yang, H.Y., Yang, J., Yang, H., Yang, G., Yao, H., Ye, X., Yin, C., Young-Moxon, B., Yu, J., Yue, X., Zhang, S., Zhang, A., Zhang, K., Zhang, Xuejie, Zhang, Lichi, Zhang, Xiaoyue, Zhang, Y., Zhang, Lei, Zhang, J., Zhang, Xiang, Zhang, T., Zhao, S., Zhao, Y., Zhao, X., Zhao, L., Zheng, Y., Zhong, L., Zhou, C., Zhou, X., Zhou, F., Zhu, H., Zhu, J., Zhuge, Y., Zong, W., Kalpathy-Cramer, J., Farahani, K., Davatzikos, C., Leemput, K.V., Menze, B., 2018. Identifying the Best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. https://doi.org/10.17863/CAM.38755.

Baum, C., Cozzo, D., Smart, N.P., 2020. Using TopGear in overdrive: a more efficient ZKPoK for SPDZ. In: Paterson, K.G., Stebila, D. (Eds.), Selected Areas in Cryptography – SAC 2019, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 274–302. https://doi.org/10.1007/978-3-030-38471-5_12.

Bendlin, R., Damgård, I., Orlandi, C., Zakarias, S., 2011. Semi-homomorphic encryption and multiparty computation. In: Paterson, K.G. (Ed.), Advances in Cryptology – EUROCRYPT 2011, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–188. https://doi.org/10.1007/978-3-642-20465-4_11.

Bhinder, B., Gilvary, C., Madhukar, N.S., Elemento, O., 2021. Artificial Intelligence in cancer research and precision medicine. Cancer Discov 11, 900–915. https://doi.org/10.1158/2159-8290.CD-21-0090.

Binder, A., Bockmayr, M., Hägele, M., Wienert, S., Heim, D., Hellweg, K., Ishii, M., Stenzinger, A., Hocke, A., Denkert, C., Müller, K.R., Klauschen, F., 2021. Morphological and molecular breast cancer profiling through explainable machine learning. Nat. Mach. Intell. 3, 355–366. https://doi.org/10.1038/s42256-021-00303-4.

Blanchard, P., El Mhamdi, E.M., Guerraoui, R., Stainer, J., et al., 2017. Machine learning with adversaries: byzantine tolerant gradient descent. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., et al. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc.

Boehm, K.M., Khosravi, P., Vanguri, R., Gao, J., Shah, S.P., 2022. Harnessing multimodal data integration to advance precision oncology. Nat. Rev. Cancer 22, 114–126. https://doi.org/10.1038/s41568-021-00408-3.

Brenner, H., Chang-Claude, J., Seiler, C.M., Stürmer, T., Hoffmeister, M., 2006. Does a negative screening colonoscopy ever need to be repeated? Gut 55, 1145–1150. https://doi.org/10.1136/gut.2005.087130.

Canetti, R., Lindell, Y., Ostrovsky, R., Sahai, A., 2002. Universally composable two-party and multi-party secure computation. In: Proceedings of the Thiry-Fourth Annual ACM Symposium on Theory of Computing. Presented at the STOC02: Symposium on the Theory of Computing, ACM. Montreal Quebec Canada, pp. 494–503. https://doi.org/10.1145/509907.509980.

Carr, P.R., Weigl, K., Edelmann, D., Jansen, L., Chang-Claude, J., Brenner, H., Hoffmeister, M., 2020. Estimation of Absolute risk of colorectal cancer based on healthy lifestyle, genetic risk, and colonoscopy status in a population-based study. Gastroenterology 159, 129–138. https://doi.org/10.1053/j.gastro.2020.03.016 e9.

Catrina, O., de Hoogh, S., 2010. Secure multiparty linear programming using fixed-point arithmetic. In: Gritzalis, D., Preneel, B., Theoharidou, M. (Eds.), Computer Security – ESORICS 2010, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 134–150. https://doi.org/10.1007/978-3-642-15497-3_9.

Cheon, J.H., Kim, A., Kim, M., Song, Y., 2017. Homomorphic encryption for arithmetic of approximate numbers. In: Takagi, T., Peyrin, T. (Eds.), Advances in Cryptology – ASIACRYPT 2017, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 409–437. https://doi.org/10.1007/978-3-319-70694-8_15.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49.

Cirillo, M.D., Abramian, D., Eklund, A., 2021. What is the best data augmentation for 3D brain tumor segmentation?. In: 2021 IEEE International Conference on Image Processing (ICIP). Presented at the 2021 IEEE International Conference on Image Processing (ICIP). Anchorage, AK, USA. IEEE, pp. 36–40. https://doi.org/10.1109/ICIP42928.2021.9506328.

Costache, A., Smart, N.P., Vivek, S., Waller, A., 2017. Fixed-Point Arithmetic in SHE Schemes. In: Avanzi, R., Heys, H. (Eds.), Selected Areas in Cryptography – SAC 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 401–422. https://doi.org/10.1007/978-3-319-69453-5_22.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat. Med. 24, 1559–1567. https://doi.org/10.1038/s41591-018-0177-5.

Culjak, I., Abram, D., Pribanic, T., Dzapo, H., Cifrek, M., 2012. A brief introduction to OpenCV. In: 2012 Proceedings of the 35th International Convention MIPRO.

Damgård, I., Keller, M., Larraia, E., Pastro, V., Scholl, P., Smart, N.P., 2013. Practical covertly secure MPC for dishonest majority Or: breaking the SPDZ limits. In: Crampton, J., Jajodia, S., Mayes, K. (Eds.), Computer Security – ESORICS 2013, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–18. https://doi.org/10.1007/978-3-642-40203-6_1.

Damgård, I., Pastro, V., Smart, N., Zakarias, S., 2012. Multiparty computation from somewhat homomorphic encryption. In: Safavi-Naini, R., Canetti, R. (Eds.), Advances in Cryptology – CRYPTO 2012, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 643–662. https://doi.org/10.1007/978-3-642-32009-5_38.

Dayan, I., Roth, H.R., Zhong, A., Harouni, A., Gentili, A., Abidin, A.Z., Liu, A., Costa, A. B., Wood, B.J., Tsai, C.S., Wang, C.H., Hsu, C.N., Lee, C.K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F.C., Lacey, G., de Antônio Corradi, G.C., Nino, G., Shin, H.H., Obinata, H., Ren, H., Crane, J.C., Tetreault, J., Guan, J., Garrett, J.W., Kaggie, J.D., Park, J.G., Dreyer, K., Juluru, K., Kersten, K., Rockenbach, M.A.B.C., Linguraru, M. G., Haider, M.A., AbdelMaseeh, M., Rieke, N., Damasceno, P.F., e Silva, P.M.C., Wang, P., Xu, S., Kawano, S., Sriswasdi, S., Park, S.Y., Grist, T.M., Buch, V., Jantarabenjakul, W., Wang, W., Tak, W.Y., Li, X., Lin, X., Kwon, Y.J., Quraini, A., Feng, A., Priest, A.N., Turkbey, B., Glicksberg, B., Bizzo, B., Kim, B.S., Tor-Díez, C., Lee, C.C., Hsu, C.J., Lin, C., Lai, C.L., Hess, C.P., Compas, C., Bhatia, D., Oermann, E. K., Leibovitz, E., Sasaki, H., Mori, H., Yang, I., Sohn, J.H., Murthy, K.N.K., Fu, L.C., de Mendonça, M.R.F., Fralick, M., Kang, M.K., Adil, M., Gangai, N., Vateekul, P., Elnajjar, P., Hickman, S., Majumdar, S., McLeod, S.L., Reed, S., Gräf, S., Harmon, S., Kodama, T., Puthanakit, T., Mazzulli, T., de Lavor, V.L., Rakvongthai, Y., Lee, Y.R., Wen, Y., Gilbert, F.J., Flores, M.G., Li, Q., 2021. Federated learning for predicting clinical outcomes in patients with COVID-19. Nat. Med. 27, 1735–1743. https://doi.org/10.1038/s41591-021-01506-3.

Dwork, C., Roth, A., 2013. The Algorithmic foundations of differential privacy. Found. Trends® Theor. Comput. Sci. 9, 211–407. https://doi.org/10.1561/0400000042.

Echle, A., Rindtorff, N.T., Brinker, T.J., Luedde, T., Pearson, A.T., Kather, J.N., 2021. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br. J. Cancer 124, 686–696. https://doi.org/10.1038/s41416-020-01122-x.

Elemento, O., Leslie, C., Lundin, J., Tourassi, G., 2021. Artificial intelligence in cancer research, diagnosis and therapy. Nat. Rev. Cancer 21, 747–752. https://doi.org/10.1038/s41568-021-00399-1.

Froelicher, D., Troncoso-Pastoriza, J.R., Pyrgelis, A., Sav, S., Sousa, J.S., Bossuat, J.P., Hubaux, J.P., 2021. Scalable privacy-preserving distributed learning.

Fu, Y., Jung, A.W., Torne, R.V., Gonzalez, S., Vöhringer, H., Shmatko, A., Yates, L.R., Jimenez-Linan, M., Moore, L., Gerstung, M., 2020. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat. Cancer 1, 800–810. https://doi.org/10.1038/s43018-020-0085-8.

GDC, n.d. URL https://portal.gdc.cancer.go.

Gentry, C., 2009. Fully homomorphic encryption using ideal lattices. In: Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. Presented at the STOC '09: Symposium on Theory of Computing. Bethesda MD USA. ACM, pp. 169–178. https://doi.org/10.1145/1536414.1536440.

Ghaffari Laleh, N., Muti, H.S., Loeffler, C.M.L., Echle, A., Saldanha, O.L., Mahmood, F., Lu, M.Y., Trautwein, C., Langer, R., Dislich, B., Buelow, R.D., Grabsch, H.I., Brenner, H., Chang-Claude, J., Alwers, E., Brinker, T.J., Khader, F., Truhn, D., Gaisa, N.T., Boor, P., Hoffmeister, M., Schulz, V., Kather, J.N., 2022. Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. Med. Image Anal. 79, 102474 https://doi.org/10.1016/j.media.2022.102474.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on image net classification. In: 2015 IEEE International Conference on Computer Vision (ICCV). Presented at the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, Santiago, Chile, pp. 1026–1034. https://doi.org/10.1109/ICCV.2015.123.

Henry, T., Carré, A., Lerousseau, M., Estienne, T., Robert, C., Paragios, N., Deutsch, E., 2021. Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a brats 2020 challenge solution. In: Crimi, A., Bakas, S. (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 327–339. https://doi.org/10.1007/978-3-030-72084-1_30.

Hutchins, G., Southward, K., Handley, K., Magill, L., Beaumont, C., Stahlschmidt, J., Richman, S., Chambers, P., Seymour, M., Kerr, D., Gray, R., Quirke, P., 2011. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 29, 1261–1270. https://doi.org/10.1200/JCO.2010.30.1366.

Ioffe, S., Szegedy, C., 2015. batch normalization: accelerating deep network training by reducing internal covariate shift.

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., Jungmann, F., Steinborn, M.M., Saleh, A., Makowski, M., Rueckert, D., Braren, R., 2021. End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nat. Mach. Intell. 3, 473–484. https://doi.org/10.1038/s42256-021-00337-8.

Kaissis, G.A., Makowski, M.R., Rückert, D., Braren, R.F., 2020. Secure, privacy-preserving and federated machine learning in medical imaging. Nat. Mach. Intell. 2, 305–311. https://doi.org/10.1038/s42256-020-0186-1.

Kather, J.N., Calderaro, J., 2020. Development of AI-based pathology biomarkers in gastrointestinal and liver cancer. Nat. Rev. Gastroenterol. Hepatol. 17, 591–592. https://doi.org/10.1038/s41575-020-0343-3.

Kather, J.N., Heij, L.R., Grabsch, H.I., Loeffler, C., Echle, A., Muti, H.S., Krause, J., Niehues, J.M., Sommer, K.A.J., Bankhead, P., Kooreman, L.F.S., Schulte, J.J., Cipriani, N.A., Buelow, R.D., Boor, P., Ortiz-Brüchle, N., Hanby, A.M., Speirs, V., Kochanny, S., Patnaik, A., Srisuwananukorn, A., Brenner, H., Hoffmeister, M., van den Brandt, P.A., Jäger, D., Trautwein, C., Pearson, A.T., Luedde, T., 2020. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat. Cancer 1, 789–799. https://doi.org/10.1038/s43018-020-0087-6.

Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., Grabsch, H.I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., Luedde, T., 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat. Med. 25, 1054–1056. https://doi.org/10.1038/s41591-019-0462-y.

Keller, M., 2020. MP-SPDZ: a versatile framework for multi-party computation. In: Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. Presented at the CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security. Virtual Event USA. ACM, pp. 1575–1590. https://doi.org/10.1145/3372297.3417872.

Keller, M., Pastro, V., Rotaru, D., 2018. Overdrive: making SPDZ Great Again. In: Nielsen, J.B., Rijmen, V. (Eds.), Advances in Cryptology – EUROCRYPT 2018, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 158–189. https://doi.org/10.1007/978-3-319-78372-7_6.

Killock, D., 2020. AI outperforms radiologists in mammographic screening. Nat. Rev. Clin. Oncol. 17 https://doi.org/10.1038/s41571-020-0329-7, 134–134.

Kingma, D.P., Ba, J., 2017. Adam: a method for stochastic optimization.

Kleppe, A., Skrede, O.J., De Raedt, S., Liestøl, K., Kerr, D.J., Danielsen, H.E., 2021. Designing deep learning studies in cancer diagnostics. Nat. Rev. Cancer 21, 199–211. https://doi.org/10.1038/s41568-020-00327-9.

Konečný, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D., 2017. Federated learning: strategies for improving communication efficiency.

Konietschke, F., Pauly, M., 2014. Bootstrapping and permuting paired t-test type statistics. Stat. Comput. 24, 283–296. https://doi.org/10.1007/s11222-012-9370-4.

Lewis, C., McQuaid, S., Clark, P., Murray, P., McGuigan, T., Greene, C., Coulter, B., Mills, K., James, J., 2018. The northern ireland Biobank: a cancer focused repository of science. Open J. Bioresour. 5, 9. https://doi.org/10.5334/ojb.47.

Li, X., Jansen, L., Chang-Claude, J., Hoffmeister, M., Brenner, H., 2022. Risk of colorectal cancer associated with lifetime excess weight. JAMA Oncol. 8, 730–737. https://doi.org/10.1001/jamaoncol.2022.0064.

Loeffler, C.M.L., Ortiz Bruechle, N., Jung, M., Seillier, L., Rose, M., Laleh, N.G., Knuechel, R., Brinker, T.J., Trautwein, C., Gaisa, N.T., Kather, J.N., 2022. Artificial intelligence-based detection of FGFR3 mutational status directly from routine histology in bladder cancer: a possible preselection for molecular testing? Eur. Urol. Focus 8, 472–479. https://doi.org/10.1016/j.euf.2021.04.007.

Lotter, W., Diab, A.R., Haslam, B., Kim, J.G., Grisot, G., Wu, E., Wu, K., Onieva, J.O., Boyer, Y., Boxerman, J.L., Wang, M., Bandler, M., Vijayaraghavan, G.R., Gregory Sorensen, A., 2021. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat. Med. 27, 244–249. https://doi.org/10.1038/s41591-020-01174-9.

Loughrey, M.B., McGrath, J., Coleman, H.G., Bankhead, P., Maxwell, P., McGready, C., Bingham, V., Humphries, M.P., Craig, S.G., McQuaid, S., Salto-Tellez, M., James, J. A., 2021. Identifying mismatch repair-deficient colon cancer: near-perfect concordance between immunohistochemistry and microsatellite instability testing in a large, population-based series. Histopathology 78, 401–413. https://doi.org/10.1111/his.14233.

Lu, M.Y., Chen, R.J., Kong, D., Lipkova, J., Singh, R., Williamson, D.F.K., Chen, T.Y., Mahmood, F., 2022. Federated learning for computational pathology on gigapixel whole slide images. Med. Image Anal. 76, 102298 https://doi.org/10.1016/j.media.2021.102298.

Lu, M.Y., Chen, T.Y., Williamson, D.F.K., Zhao, M., Shady, M., Lipkova, J., Mahmood, F., 2021. AI-based pathology predicts origins for cancers of unknown primary. Nature 594, 106–110. https://doi.org/10.1038/s41586-021-03512-4.

Ma, J., Naas, S., Sigg, S., Lyu, X., 2022a. Privacy-preserving federated learning based on multi-key homomorphic encryption. Int. J. Intell. Syst. 37, 5880–5901. https://doi.org/10.1002/int.22818.

Ma, X., Zhou, Y., Wang, L., Miao, M., 2022b. Privacy-preserving Byzantine-robust federated learning. Comput. Stand. Interfaces 80, 103561. https://doi.org/10.1016/j.csi.2021.103561.

Macenko, M., Niethammer, M., Marron, J.S., Borland, D., Woosley, J.T., Guan, Xiaojun, Schmitt, C., Thomas, N.E., 2009. A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Presented at the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI). Boston, MA, USA. IEEE, pp. 1107–1110. https://doi.org/10.1109/ISBI.2009.5193250.

Marks, K., West, N., 2020. Molecular assessment of colorectal cancer through Lynch syndrome screening. Diagn. Histopathol. 26, 47–50. https://doi.org/10.1016/j.mpdhp.2019.10.012.

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., De Fauw, J., Shetty, S., 2020. International evaluation of an AI system for breast cancer screening. Nature 577, 89–94. https://doi.org/10.1038/s41586-019-1799-6.

McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.y, 2017. Communication-efficient learning of deep networks from decentralized data.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Criminisi, A., Das, T., Delingette, H., Demiralp, C., Durst, C.R., Dojat, M., Doyle, S., Festa, J., Forbes, F., Geremia, E., Glocker, B., Golland, P., Guo, X., Hamamci, A., Iftekharuddin, K.M., Jena, R., John, N.M., Konukoglu, E., Lashkari, D., Mariz, J.A., Meier, R., Pereira, S., Precup, D., Price, S.J., Raviv, T.R., Reza, S.M.S., Ryan, M., Sarikaya, D., Schwartz, L., Shin, H.C., Shotton, J., Silva, C.A., Sousa, N., Subbanna, N.K., Szekely, G., Taylor, T.J., Thomas, O.M., Tustison, N.J., Unal, G., Vasseur, F., Wintermark, M., Ye, D.H., Zhao, L., Zhao, B., Zikic, D., Prastawa, M., Reyes, M., Van Leemput, K., 2015. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging 34, 1993–2024. https://doi.org/10.1109/TMI.2014.2377694.

Quasar Collaborative Group, Gray, R., Barnwell, J., McConkey, C., Hills, R.K., Williams, N.S., Kerr, D.J., 2007. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. Lancet Lond. Engl. 370, 2020–2029. https://doi.org/10.1016/S0140-6736(07)61866-2.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

Saldanha, O.L., Quirke, P., West, N.P., James, J.A., Loughrey, M.B., Grabsch, H.I., Salto-Tellez, M., Alwers, E., Cifci, D., Ghaffari Laleh, N., Seibel, T., Gray, R., Hutchins, G.G. A., Brenner, H., van Treeck, M., Yuan, T., Brinker, T.J., Chang-Claude, J., Khader, F., Schuppert, A., Luedde, T., Trautwein, C., Muti, H.S., Foersch, S., Hoffmeister, M., Truhn, D., Kather, J.N., 2022. Swarm learning for decentralized artificial intelligence in cancer histopathology. Nat. Med. 28, 1232–1239. https://doi.org/10.1038/s41591-022-01768-5.

Sav, S., Pyrgelis, A., Troncoso-Pastoriza, J.R., Froelicher, D., Bossuat, J.P., Sousa, J.S., Hubaux, J.P., 2021. POSEIDON: privacy-preserving federated neural network learning.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591. https://doi.org/10.2307/2333709.

Stripelis, D., Saleem, H., Ghai, T., Dhinagar, N.J., Gupta, U., Anastasiou, C., Ver Steeg, G., Ravi, S., Naveed, M., Thompson, P.M., Ambite, J.L., 2021. Secure neuroimaging analysis using federated learning with homomorphic encryption. In: Walker, A., Rittner, L., Romero Castro, E., Lepore, N., Brieva, J., Linguraru, M.G. (Eds.), 17th International Symposium on Medical Information Processing and Analysis. Presented at the Seventeenth International Symposium on Medical Information Processing and Analysis. Campinas, Brazil. SPIE, p. 44. https://doi.org/10.1117/12.2606256.

Taiello, R., Önen, M., Humbert, O., Lorenzi, M., 2022. Privacy Preserving Image Registration. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, Lecture Notes in Computer Science. Springer Nature Switzerland, Cham, pp. 130–140. https://doi.org/10.1007/978-3-031-16446-0_13.

Tayebi Arasteh, S., Ziller, A., Kuhl, C., Makowski, M., Nebelung, S., Braren, R., Rueckert, D., Truhn, D., Kaissis, G., 2023. Private, fair and accurate: training large-scale, privacy-preserving AI models in medical imaging.

Taylor, J., Wright, P., Rossington, H., Mara, J., Glover, A., West, N., Morris, E., Quirke, P., YCR BCIP study group, 2019. Regional multidisciplinary team intervention programme to improve colorectal cancer outcomes: study protocol for the Yorkshire Cancer Research Bowel Cancer Improvement Programme (YCR BCIP). BMJ Open 9, e030618. https://doi.org/10.1136/bmjopen-2019-030618.

Truex, S., Baracaldo, N., Anwar, A., Steinke, T., Ludwig, H., Zhang, R., Zhou, Y., 2019. A hybrid approach to privacy-preserving federated learning. In: Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security. Presented at the CCS '19: 2019 ACM SIGSAC Conference on Computer and Communications Security. London United Kingdom. ACM, pp. 1–11. https://doi.org/10.1145/3338501.3357370.

Usynin, D., Ziller, A., Makowski, M., Braren, R., Rueckert, D., Glocker, B., Kaissis, G., Passerat-Palmbach, J., 2021. Adversarial interference and its mitigations in privacy-preserving collaborative machine learning. Nat. Mach. Intell. 3, 749–758. https://doi.org/10.1038/s42256-021-00390-3.

Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., Qi, H., 2019. Beyond Inferring class representatives: user-level privacy leakage from federated learning. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Presented at the IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Paris, France. IEEE, pp. 2512–2520. https://doi.org/10.1109/INFOCOM.2019.8737416.

Warnat-Herresthal, S., Schultze, H., Shastry, K.L., Manamohan, S., Mukherjee, Saikat, Garg, V., Sarveswara, R., Händler, K., Pickkers, P., Aziz, N.A., Ktena, S., Tran, F., Bitzer, M., Ossowski, S., Casadei, N., Herr, C., Petersheim, D., Behrends, U., Kern, F., Fehlmann, T., Schommers, P., Lehmann, C., Augustin, M., Rybniker, J., Altmüller, J., Mishra, N., Bernardes, J.P., Krämer, B., Bonaguro, L., Schulte-Schrepping, J., De Domenico, E., Siever, C., Kraut, M., Desai, M., Monnet, B., Saridaki, M., Siegel, C.M., Drews, A., Nuesch-Germano, M., Theis, H., Heyckendorf, J., Schreiber, S., Kim-Hellmuth, S., COVID-19 Aachen Study (COVAS), Nattermann, J., Skowasch, D., Kurth, I., Keller, A., Bals, R., Nürnberg, P., Rieß, O., Rosenstiel, P., Netea, M.G., Theis, F., Mukherjee, Sach, Backes, M., Aschenbrenner, A.C., Ulas, T., Deutsche COVID-19 Omics Initiative (DeCOI), Breteler, M.M.B., Giamarellos-Bourboulis, E.J., Kox, M., Becker, M., Cheran, S., Woodacre, M.S., Goh, E.L., Schultze, J.L., 2021. Swarm learning for decentralized and confidential clinical machine learning. Nature 594, 265–270. https://doi.org/10.1038/s41586-021-03583-3.

Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R.M., Rubin, D.L., Lungren, M.P., 2020. Preparing medical imaging data for machine learning. Radiology 295, 4–15. https://doi.org/10.1148/radiol.2020192224.

Zhang, C., Li, S., Xia, J., Wang, W., Yan, F., Liu, Y., 2020. BatchCrypt: efficient homomorphic encryption for cross-silo federated learning. In: Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC 2020), pp. 493–506.

Zhao, B., Mopuri, K.R., Bilen, H., 2020. iDLG: improved deep leakage from gradients.

Ziller, A., Mueller, T.T., Braren, R., Rueckert, D., Kaissis, G., 2022. Privacy: an axiomatic approach. Entropy 24, 714. https://doi.org/10.3390/e24050714.

Ziller, A., Trask, A., Lopardo, A., Szymkow, B., Wagner, B., Bluemke, E., Nounahon, J.M., Passerat-Palmbach, J., Prakash, K., Rose, N., Ryffel, T., Reza, Z.N., Kaissis, G., 2021. PySyft: a Library for easy federated learning. In: Rehman, M.H.ur, Gaber, M.M. (Eds.), Federated Learning Systems, Studies in Computational Intelligence. Springer International Publishing, Cham, pp. 111–139. https://doi.org/10.1007/978-3-030-70604-3_5.