Privacy Risks of Whole-Slide Image Sharing in Digital Pathology

Petr Holub[*1,2], Heimo Müller[3], Tomáš Bíl[2], Luca Pireddu[5], Markus Plass[3], Fabian Prasser[6], Irene Schlünder[7], Kurt Zatloukal[3], Rudolf Nenutil[8], and Tomáš Brázdil[2,4]

[1] *BBMRI-ERIC, Graz, Austria*
[2] *Institute of Computer Science, Masaryk University, Brno, Czech Republic*
[3] *BBMRI.at & Diagnostic & Research Center for Molecular BioMedicine, Medical University of Graz, Graz, A-8010, Austria*
[4] *Faculty of Informatics, Masaryk University, Brno, Czech Republic*
[5] *Visual and Data-intensive Computing Group, CRS4, Pula, Italy*
[6] *Berlin Institute of Health @ Charité – Universitätsmedizin Berlin, Berlin, Germany*
[7] *TMF eV, Berlin, Germany*
[8] *BBMRI.cz & Masaryk Memorial Cancer Institute, Czech Republic*
[*] *Corresponding author: Assoc. Prof. Petr Holub, Ph.D., BBMRI-ERIC, Neue Stiftingalstraße 2/B/6. 8010 Graz, Austria.* <petr.holub@bbmri-eric.eu>

April 25, 2022

Abstract

Access to large volumes of so called whole-slide images, high-resolution scans of complete pathological slides, has become a cornerstone of development of novel artificial intelligence methods in digital pathology, but has broader impact for medical research and education/training. However, a methodology based on risk analysis for sharing such imaging data and applying the principle "as open as possible and as closed as necessary" is still lacking. In this article we develop a model for privacy risk analysis for whole-slide images, which focuses primarily on identity disclosure attacks, as these are the most important from a regulatory perspective. We develop a mathematical model for risk assessment and design a taxonomy of whole-slide images with respect to privacy risks. Based on these risk assessment model and the taxonomy, we design a series of experiments to demonstrate the risks on real-world imaging data. Finally, we develop guidelines for risk assessment and recommendations for low-risk sharing of whole-slide image data.

1 Introduction and Motivation

The last decade has seen tremendous advances in the methods available to pathologists for computer-assisted diagnosis, particularly thanks to the rapid developments in digital microscopy,

which has reached high interchangeability levels with optical microscopy [1]; the whole field has become known as *digital pathology*. [2] Availability of large volumes of imaging and other types of clinically relevant data and availability of large-scale compute capacities has resulted in the massive development of Artificial Intelligence (AI) methods aiming to support pathologists in the diagnostic process.[3]

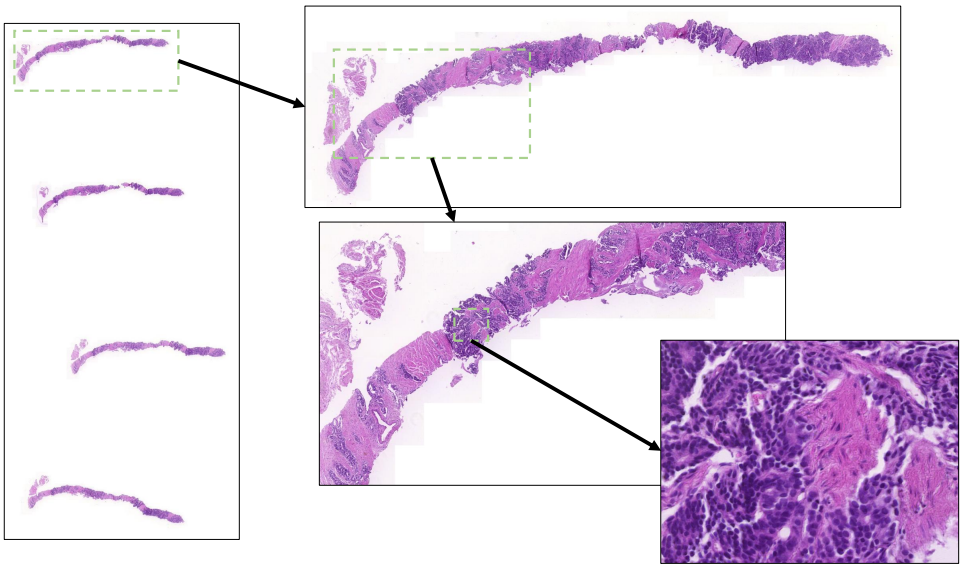
The fundamental data used in this domain are whole-slide images (WSIs): high-resolution optical microscopy scans of the whole slide of the biological material, resulting in image data typically in the order of gigapixels or tens of gigapixels, as shown in Figure 1. WSIs are widely used for purposes ranging from routine diagnostics to development and application of AI models. The images are commonly stored in databases linked to other types of the data—e.g., as a part of hospital information systems—and they are sometimes even shared under tight confidentiality agreements (e.g., ADOPT CRC-Cohort[1]) or under the assumption of inherent anonymity as open data (e.g., CAMELYON competition [4, 5] or TCGA Digital Slide Archive [6]).

The process of obtaining a WSI starts with acquisition of the biological material from a patient in a surgery or a biopsy. The material is then cut into blocks that are formalin-fixed and paraffin-embedded (hence the FFPE abbreviation), which are then sectioned and mounted onto glass slides and stained (colored) based on the type of the material and diagnostic methods to be applied—most common stainings being hematoxylin-eosin, van Gieson or various modern immunochemical staining methods. The material is then digitized using an slide scanner in the visible or fluorescence spectrum at $<0.250\,\mu m/px$ (usually designated as 20×) or $<0.125\,\mu m/px$ (usually denoted 40×). The resulting images are typically in the size range of gigapixels or tens of gigapixels and show detailed and irregular cellular structure of the tissue, as shown in Figure 1. In some cases the scan can also include a visible patient identifier, e.g., a bar code that could be a patient-related ID or or a research code (pseudonym). Metadata in the resulting file usually include details about the scanner and its settings.

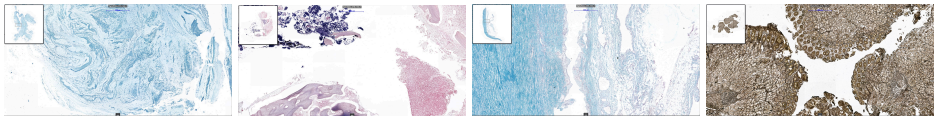
Given the large amounts of data required for the development of AI models, it is necessary to access large WSI collections, or even assemble collections by pooling data from different sources. However, the privacy risks related to WSI sharing have not yet been systematically explored and the practice of sharing is extremely heterogeneous: from the above-mentioned approaches considering the WSI low-risk data and sharing them as open datasets, to considering them as sensitive as other clinical data and sharing them only as a part of pseudonymized data sets under contracts compliant relevant data protection laws—e.g., General Data Protection Regulation (GDPR) in European countries.

Hence, the pivotal question is: *what are the privacy risks related to sharing WSIs and are there any circumstances under which the risks can be considered low enough to treat the data as anonymous with relatively low privacy risks?* At a first glance it would not appear feasible to identify the source individual from a WSI image content, as the biological material has already been removed from the patient's body and therefore histological slides and WSIs produced from this tissue cannot be directly linked to a person. On the other hand, the WSIs are very large with relatively characteristic tissue structure, making each image very unique and potentially enabling the extraction of identifying information. There are also potential artifacts, such as

[1] https://www.bbmri-eric.eu/scientific-collaboration/colorectal-cancer-cohort/



(a)



(b)

Figure 1: Examples of WSIs and details of its resolution. (a) Prostate cancer biopsy image has been stained with hematoxylin-eosin staining and scanned at 20×, resulting in resolution of approximately 100,000 px × 200,000 px. (b) Various types of less common staining methods, left to right: Giemsa, Gram, Alcian Blue Stain, and Warthin–Starry (silver) stain.

slide preparation characteristics, associated metadata and various other factors that can significantly increase the likelihood with which a WSI can be traced back to its original slide and associated data.

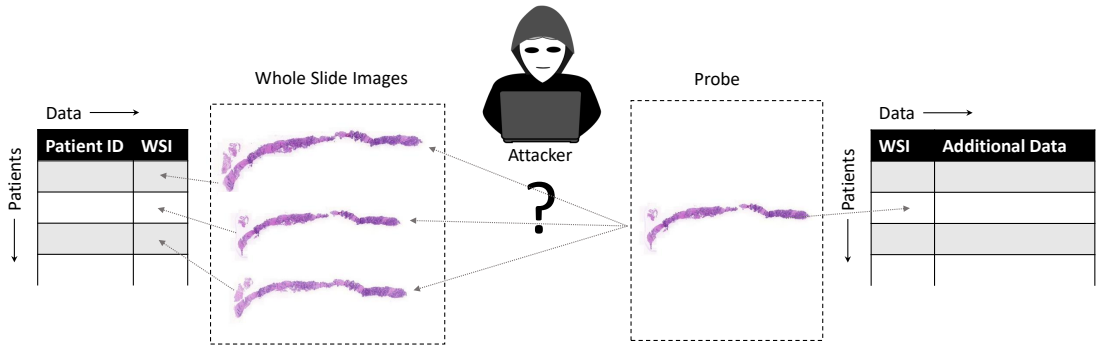


Figure 2: Assumed model for data structure—tabular data with patients in rows and different data types in columns, demonstrating that different data sets can be linked by a WSI.

Privacy risks can be understood in two orthogonal dimensions: *likelihood* and resulting *harm* of a successful attack. In this paper, we focus on analysis of likelihood dimension for linking disclosure attacks, where the WSIs can act as a key for linking different data sets. This can lead to identity disclosure or to inferring additional information about data subjects contributing to the data sets containing WSI data. The level of harm is then dependent on what specific additional information is available and we consider it fixed for simplicity; in practical risk assessment it comes as an additional input.

We consider the following problem, illustrated in Figure 2. Assume that an attacker possesses a *background knowledge*, and a dataset of patients indexed by a *PatientID* (for simplicity we assume that PatientID is unique), containing WSI data linked to patients. Now the attacker is given another WSI, called a *probe*, with some *additional data* attached. We assume that the probe belongs to one of the patients in the dataset, but does necessarily appear in the dataset (may be, e.g., a different WSI of slide cut from the same block of material of the patient). The aim of the attacker is to link the probe to the correct patient, by assigning the probe to the WSI in the background knowledge, and hence also link additional data to the correct patient. In case the attack is successful, the attacker expands his knowledge.

We consider the following crucial questions.

- How to quantitatively measure chances of attacker's success? We design a metric, vaguely inspired by differential privacy (but not equivalent), which measures the probability of successful identification for deterministic or randomized attacks.

- What WSI can be used by the attacker as a potential efficient probe? We introduce a hierarchical taxonomy capturing "closeness" of different WSIs with respect to their linking potential, starting with identical WSIs, progressing through different scans of the same slide and WSIs coming from tissues with different degree of spatial or temporal closeness.

Based on the introduced metric and the slide hierarchy, we perform a series of experimental evaluations using state-of-the-art image similarity techniques, to demonstrate that the attacker's chances of success depend on the slide's location in the taxonomy. The experimental evaluation utilizes the extensive WSI collections of the Medical University Graz and the Masaryk Memorial Cancer Institute to evaluate likelihood of a successful attack in real-world settings.

We outline guidelines for the conscientious sharing of WSI data, encouraging the advancement of data-driven research while protecting the identity of the individuals from whom the tissue was originally extracted. Our risk assessment methodology also provides practical approaches for data controllers to perform related analyses for their own data sets.

The paper is structured as follows. Section 2 describes the privacy risk analysis approach adopted in the paper, rigorously defines the attack model so that it can be measured in experiments, defines the taxonomy of the WSI with respect to privacy risks, and provides an overview of the attack experiments and datasets used for them. Results of the experiments are summarized and discussed in Section 3. Based on these results, we propose guidelines to support sharing of WSI data in Section 4 and also discuss possible future privacy attack aspects that can be further explored.

2 Methods

2.1 Characterization of privacy risks considered

In our model we consider privacy risks caused by WSIs acting as (a) direct or (b) indirect links across different data. Privacy attacks can thus occur either by using the WSIs directly, or by using any other data available in the data set when WSIs are used to identify accurately or approximately a data subject. In this section we introduce person identification and information gain as two independent dimensions, which we further use for modeling attacks.

The attack model assumes that an attacker that has a background knowledge, including WSI data, receives a WSI probe with additional data attached. The attacker tries to correctly assign this probe to the patients in the background knowledge. This operation can be performed deterministically—e.g., based on image similarity. To obtain a general model, we consider potentially randomized attackers who may guess the assignment using, e.g., a series of coin tosses in addition to the available information. Note that the deterministic assignment is a special case of the randomized one where each probe is assigned to a single patient with probability one.

The intuition behind the metric. Given a model of an attacker f, we define the *attack value* $V_{\text{att}}(f)$ to be the proportion of the patients in the attacker's background knowledge that are correctly guessed from available probes. This means the proportion of patients assigned by f to at least one of their probes and thus potentially compromised.

Formal metric for deterministic attacks. Consider a set $\mathcal{H}$ of *patients* and a set $\mathcal{P}$ of data *probes* that the attacker is trying to map to the patients in $\mathcal{H}$. Given a patient $h \in \mathcal{H}$, the attacker may acquire a probe belonging to the patient. Formally, we define the *probe ground*

truth as a function which assigns a non-empty set of possible probes $G(h)$ to every patient h. Formally G is defined as $G : \mathcal{H} \to 2^{\mathcal{P}}$, where $2^{\mathcal{P}}$ is the set of all subsets, so called powerset, of $\mathcal{P}$. In our case $G(h)$ consists of all publicly available WSIs of the patient h. The probe ground truth is not known to the attacker.

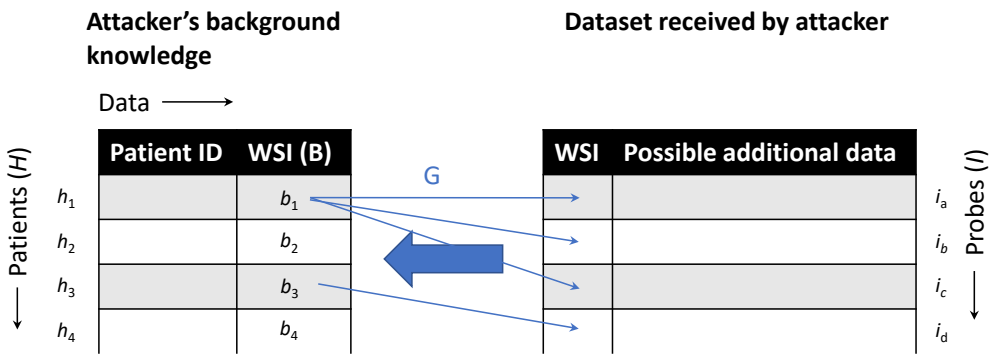


Figure 3: Illustration of G. Example G defines which probes belong to which patients according to the "ground truth" (not known to the attacker): $G(h_1) = \{i_a, i_b, i_c\}$ and $G(h_2) = \{i_d\}$.

Consider an *attack* $f : \mathcal{P} \to \mathcal{H}$ on the identity of the patient h using the probe p. That is, given a probe p, the attack f assigns a patient $f(p) \in \mathcal{H}$ to the probe p. We say that a given patient $h \in \mathcal{H}$ is *f-vulnerable* if there is a probe $p \in G(h)$ such that $h = f(p)$; that is, if at least one of the probes of the patient h is correctly assigned to h.

We define *worst-case probe attack value* by

$$V_{\text{att}}(f) = \frac{\text{number of } f\text{-vulnerable patients}}{|\mathcal{H}|}$$

i.e., ratio of number of f-vulnerable patients to total number of patients.

Observe that there is always a perfect attack f under which all patients are f-vulnerable: a perfect attack simply assigns correctly a patient to at least one of his/her probes. However, such an attack cannot be practically implemented as attackers typically have limited background knowledge. So we consider families of possible attacks (*attack domains*) based on the type of the background knowledge and algorithms used for the probe-patient assignment. In this work we consider attack f that map probes to patients according to the WSI similarity. However, the above defined metric can be understood in a more general sense as a method of measuring an attack on partially anonymized datasets using publicly available data.

Formally, we say that the set of patients $\mathcal{H}$ and a set of probes $\mathcal{P}$ is *ε-safe with respect to an attack domain* $\mathcal{F}$, if for every $f \in \mathcal{F}$ we have that $V_{\text{att}}(f) \leq \varepsilon$.

Application to WSI linkability risks. In this paper we apply the metric to the analysis of linking risks related to WSI data. The attacks $f \in \mathcal{F}$ are implemented by various types of algorithms that are able to deterministically assign WSI probes to patients with WSIs in the attacker's background knowledge utilizing image similarity (attack domain $\mathcal{F}$). The analysis

assumes the attacker does not have any explicit information about the mapping of the probes $\mathcal{P}$ to patients in $\mathcal{H}$. The resulting V_{att} is a proportion of patients for which given attack f is able to link WSI from the probes to the WSI from background knowledge. The set of patients $\mathcal{H}$ and the set of probes $\mathcal{P}$ are ε-safe, if for all the attacks tested ($\mathcal{F}$), the V_{att} is smaller than ε. Note that it is usually intractable to *prove* ε-safety with respect to possible attacks from a reasonably large domain. However, in some cases we show that data is *not* ε-safe with respect to relatively large ε, which means that there is an attack that correctly assigns a relatively large proportion of patients to their probes.

Note that an attack $f : \mathcal{P} \rightarrow \mathcal{H}$ is determined by three components: set of patients in attacker's background knowledge $\mathcal{H}$, set of probes $\mathcal{P}$, and set of algorithms implementing the assignment of probes to patients. From a practical perspective, in order to analyze the risks we need to establish relationship between WSIs which might appear in the attacker's background knowledge and in set of probes.

Now note given the large size and detailed tissue structure of WSIs, it is reasonable to expect that if an identical WSI appears in both sets, a trivial bit-by-bit comparison algorithm would be able to assign this probe to the correct patient deterministically and unambiguously. Comparing larger data sets can be made more efficient by using cryptographic hash functions on each WSI and comparing resulting hashes. But to what extent does this apply also to other WSIs related by spatial or temporal relation to a single patient? In order to have basis for such analysis, we develop a taxonomy of WSI relations in the next section.

2.2 WSI hierarchy

To analyze WSI linking attacks, we need to examine the factors in the image generation process that can make two WSI (e.g., one in the attacker's background knowledge and one a probe) more or less easily linked to each other. The first factor is when and from where the tissue on the slide was extracted. We introduce the following *spatiotemporal hierarchy* of cases, ordered by decreasing potential image similarity:

(H-1) WSIs of the same slide *scanned with different parameters* (including age of the slide and time of scanning, scanning parameters such as resolution, type of scanner, etc.; see Figure 4);

(H-2) WSIs of from the *same tissue block*—which can be further divided into consecutive (adjacent, see Figure 5) vs. distant slides;

(H-3) WSIs from the *same primary sample*—which can be further divided into multiple different blocks;

(H-4) WSIs from *different primary samples* (such as the primary tumor site, the lymph node, the metastasis) from the same patient *taken at the same time*;

(H-5) WSIs from different primary samples from the same patient *taken at different times*, which can also relate different diagnoses of the same patient.

Figure 4: Example of the same colon tissue from MIDI and FLASH scanners (scanners specified in detail in Section 2.3.2).

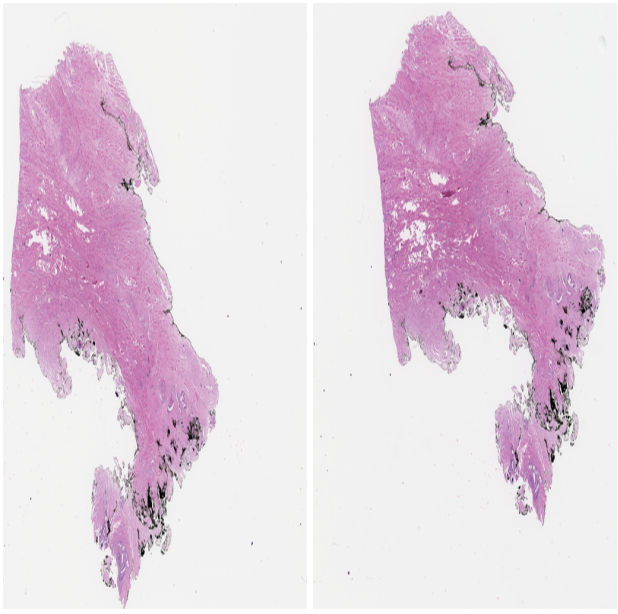


Figure 5: Example of the consecutive slides

The second factor in the image generation process affecting image similarity is the staining applied in slide preparation. We introduce the following hierarchy of *staining* cases, ordered by decreasing potential image similarity, which can be combined with the spatiotemporal hierarchy to reason about risks:

(H-a) WSI slides from *same staining batch*;

(H-b) WSIs *different staining batches using the same staining method*;

(H-c) WSIs from slides stained *using different stainings* (e.g., H&E vs. DAB-based immunohistochemistry vs. Van Glieson).

Furthermore, when considering WSIs from the privacy risk perspective, it is useful to consider them in terms of the conditions required for any variable to be considered an identifier [7]: distinguishability, replicability, and availability.

Distinguishability refers to the ability of the data to act as a fingerprint distinguishing the donor. One WSI is always distinguishing using direct bit-by-bit comparison; hence when the same WSI appears in two data sets, they are directly linkable. For different WSIs with increasing distance in the hierarchy (H-1)→(H-5) and to lesser extent also (H-a)→(H-c), each of them is still individually distinguishing, but their linkability is likely to decrease as demonstrated in the experimental part of this paper.

Replicability refers to what extent one can reproduce the data—either in succession or, more importantly, in more distant points in time. For WSIs the replicability is likely to decrease rapidly with increasing distance in spatio-temporal hierarchy. If the same slide is physically available (H-1), it can be scanned again, resulting in a WSI very similar to the previously obtained WSI from the same slide, with small differences related to aging of the slide (fading of staining), differences in physical preparation for scanning (e.g., dust particles), and properties of the scanner. Other means to replicability imply producing multiple very similar slides from different biological material. If different slides from the same block of tissue are used (H-2), the biological structures are not identical and the near perfect replicability is impossible. However, even imperfect replicability can still lead to linkability, as demonstrated later in this paper. For more distant temporal relation (H-5), replicability is unlikely due to the previous material being fixated in the slide preparation process, while new biological material remaining in the patient's body is subject to further biological development (e.g., natural growth, shrinkage/dying as reaction to treatment), making new samples diverge from previous ones. In addition, regarding tumor tissue, this is intentionally removed during surgery, hence new relevant material is no longer available in the patient's body subsequent slide replication.

Availability (also sometimes denoted as "knowability" in European Medical Agency guidelines[2]) of WSI data has been evolving recently: a decade ago, the WSI was mostly limited to diagnostic purposes, if available as digital imaging at all. Some digital imaging has been used for education and training of medical professionals [8–10], but not to the extent of substantially increasing likelihood of identification except for diseases that are so rare that they already identify the patient. The above mentioned CAMELYON competitions demonstrated,

[2] https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-1.pdf

however, how the situation is changing due to the demands and hopes put into development of novel AI methods to support cancer diagnosis and treatment; data are becoming rapidly available either under contracts or even as open data sets, thus substantially increasing availability. More recently, some publicly available data sets such as PCam [11] have been made available as (labeled) tiles, instead of full WSI. This was done primarily to simplify training in tile-based AI models, but as a side effect it can also decrease linking risks if only limited number of tiles is released from each contributing WSI. Tile-based approaches are also considered in the experimental evaluation below.

2.3 Design of Experimental Evaluation of WSIs Linking Risks

2.3.1 The attacker model implementation

We consider attackers assigning WSI probes to a patients' WSIs using similarity measures on features extracted from WSIs using common deep learning methods. Specifically, assume that the attacker is given a WSI probe p (i.e., probes consisting of WSI data only) and a background knowledge of WSIs $b_1, \dots, b_n \in \mathcal{B}$ associated the patients $h_1, \dots, h_n$, respectively. Then the attacker proceeds as follows:

- Apply *feature extractor* to all WSIs, obtaining feature vectors $w[p], w[b_1], \dots, w[b_n] \in \mathbb{R}^k$ corresponding to the WSIs. Here k is the number of features extracted from each WSI and is typically in thousands.
- Apply *similarity measure* M to all pairs of vectors $w[p]$ and $w[b_i]$, obtaining $M(w[p], w[b_i])$ for all $1 \leq i \leq n$.
- Assign the WSI probe p to the patient with the WSI b_i with the maximum $M(w[p], w[b_j])$ among all $b_1, \dots, b_n$, i.e.,

$$M(w[p], w[b_i]) = \max_j \left\{ M(w[p], w[b_j]) \mid j = 1, \dots, n \right\}$$

Feature extraction. The purpose of the feature extraction is to squeeze the high-dimensional WSIs into numerical vectors of smaller dimension. A conventional approach is to use a trained deep neural network model which takes images as inputs and outputs their vector representations, while preserving important features of the WSIs. To implement the feature extraction, we consider the following neural network models pre-trained on the ImageNet dataset [12]:

- ResNet [13] – output dimension 2,048 features;
- VGG16 [14] – output dimension 25,088 features;
- Inception – output dimension 51,200 features;
- Img2vec[3] – a python package that uses a pre-trained ResNet [15] network to extract features of the dimension 512;

[3] `https://pypi.org/project/img2vec-pytorch/`

- SimCLRv2 [16] – fine-tuned model on 100% of labels with output dimension of 8,192 features;

In each of these cases we remove the top layers of these networks, that originally solve image recognition problems, and use their internal representations of the input WSIs as feature vectors. Note that each WSI needs to be either downscaled or cropped to fit as an input of these networks, which is 224 px × 224 px. In our analysis we consider both alternatives and compare attacks using complete downscaled WSIs with attacks using WSIs cropped to their central parts.

Similarity measures. To measure similarity of feature vectors we use the following two standard measures:

- *Cosine similarity*: Given two vectors $v = (v_1, \dots, v_n)$ and $w = (w_1, \dots, w_n)$ the cosine similarity is defined by

$$M_{\cos}(v, w) = \frac{\sum_{i=1}^{n} v_i w_i}{\sqrt{\sum_{i=1}^{n} v_i^2}\sqrt{\sum_{i=1}^{n} w_i^2}}$$

- *Euclidean similarity*: Given two vectors $v = (v_1, \dots, v_n)$ and $w = (w_1, \dots, w_n)$ the Euclidean similarity is defined by

$$M_{\text{euc}}(v, w) = \frac{1}{\sqrt{\sum_{i=1}^{n} (v_i - w_i)^2}}$$

Note that $M_{\cos}$is normalized in the range $[-1; 1]$, while $M_{\text{euc}}$is not normalized.

2.3.2 Datasets

Slides for the (H-1) hierarchy analysis were provided by Masaryk Memorial Cancer Institute (MMCI) and contained colorectal cancer WSIs, breast cancer WSIs, and prostate biopsies. Slides for all the other hierarchy analyses were prostate biopsies provided by Medical University Graz (MUG). All images were stained using hematoxyline & eosin staining in a single batch; therefore, the experimental analyses are restricted to the (H-a) step from the WSI staining hierarchy.

WSI linking attacks. The attack analyses focused primarily on linking complete WSIs and we performed experiments with the following datasets based on the WSI spatiotemporal hierarchy.

(E-1) [Based on (H-1)] Consists of 28 slides from different patients provided by MMCI, which were scanned on Pannoramic® MIDI and Pannoramic® 250 Flash III scanners by 3DHistech (Budapest, Hungary) at a resolution of 0.172 µm/px. Of each pair, one WSI is randomly assigned to the probe set $\mathcal{P}$ and the other is assigned to $\mathcal{B}$ of background knowledge.

(E-2) [Based on (H-2)] Consists of WSIs from consecutive and non-consecutive prostate biopsy slides provided by MUG, which were scanned on Aperio AT2 scanner at a resolution of 0.25 µm/px. Note that in the following text we use "slides" for readability reasons, but precisely speaking the sets $\mathcal{B}$ and $\mathcal{P}$ are populated by the WSIs from the discussed slides.

(E-2a) *Consecutive slides.* The whole dataset consists of 158 pairs of (directly) consecutive WSIs (see Figure 5). Pairs are visually similar, but there are also visible differences. One randomly selected slide from each pair goes into $\mathcal{B}$—the attacker's background knowledge—and the other into the probe set $\mathcal{P}$.

(E-2b) *Non-consecutive slides.* Consists of 558 slides from 80 patients (average 6.975 slides per patient). Metadata on these slides contained information on order of cuts and approximate distances between them (ranging 3 mm to 5 mm). We used this information to compose various data subsets to study the influence of the distance between slides in background knowledge and probes (see Figure 6). For each threshold distance l and each patient, the sets $\mathcal{B}$ and $\mathcal{P}$ were populated as follows: a pivot slide is selected randomly from the patient and is inserted into $\mathcal{B}$; then, all the patient's slides that are at a distance greater than l from the pivot are added to the $\mathcal{P}$. Note that the set of slides added to $\mathcal{P}$ may be empty for some patients, if there are no slides available in distance greater than l; this practically happens for larger threshold distances. In such a case, the pivot slide in $\mathcal{B}$ for the given patient has no corresponding slides in $\mathcal{P}$.

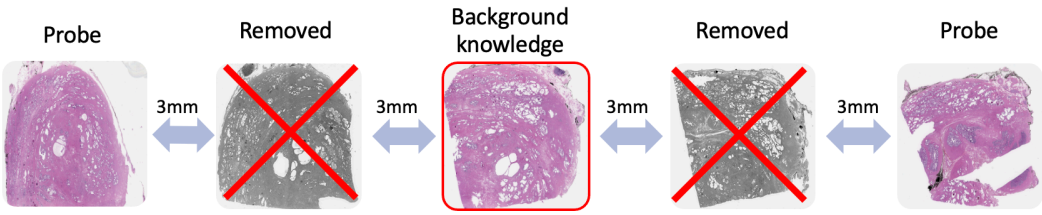
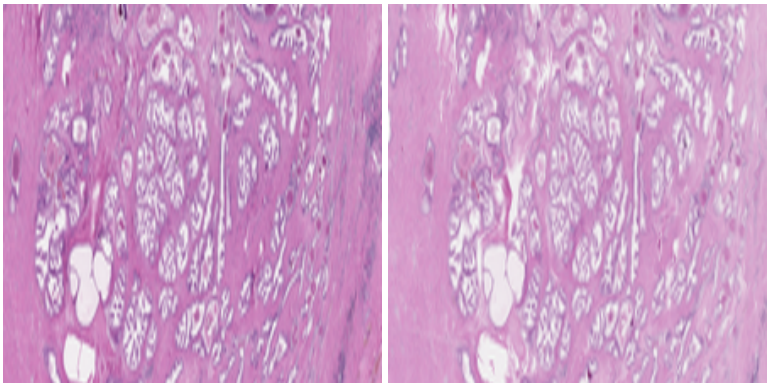


Figure 6: Example of probe selection for distance threshold l 4mm from the pivot slide selected as background knowledge

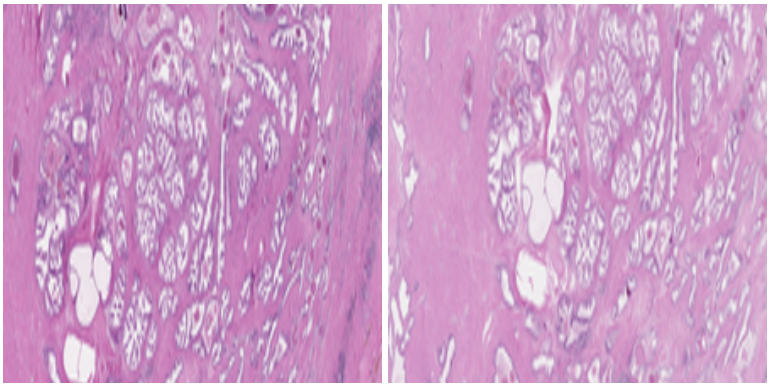
Hierarchy elements (H-4) and (H-5) were not experimentally evaluated as the methods tested already showed a dramatic decrease in V_{att} even in (H-2) with increasing spatial distance of slides from the same primary sample.

Cropped WSI linking attacks. We study also ability to attack the dataset when the overall shape of the tissue cannot contribute to the extracted features and hence to similarity. We cropped the WSIs so that the resulting image contained internal parts of the tissue only (further denoted as cropped WSIs). Registration (alignment) of the WSIs was done before the cropping, to model worst case scenario that the attacker can get access to spatially corresponding cropped WSIs (see Figure 7a). We studied influence of decreasing overlap of these cropping regions on V_{att} too, by shifting the crop regions by fixed amounts in random direction for each WSI

pair (see Figure 7b). These experiments were done for both (cE-2a) consecutive slides and (cE-2b) non-consecutive slides.



(a)



(b)

Figure 7: Examples of consecutive slides used for cropped WSI linking attacks (a) without shift and (b) with shift of 50 px. Area overlap for 50 px is 71% to 78% depending on the shift direction. Note that the images are shown already downscaled to fit input of the network as described in Section 2.3.1.

3 Results and Discussion

In this section we present the evaluation of the V_{att} value for the various WSI linking attacks described in Section 2.3.2. We have tested two similarity metrics, cosine and Euclidean, and the feature extraction using deep learning models listed in Section 2.3.1. Experiments were repeated 20 times, with all datasets described in Section 2.3.2 being regenerated randomly from the available data for each run. Naturally, at each run we measure a different V_{att} value, and the results presented in the tables below are the average V_{att} value computed from the 20 measurements.

3.1 Results on WSI linking attacks

Table 1 presents the average V_{att} value measured in the experiment (E-1), which measures the effect of the different scanning conditions on V_{att}. The results show that WSIs of the same tissue acquired using different scanners are very similar and can be used in a successful WSI linking attack.

Metric	ResNet	VGG16	Img2vec	Inception	Autoencoder	SimCLRv2
Cosine	0.951	1.0	0.951	0.980	0.757	1.0
Euclidean	0.951	1.0	0.955	0.901	0.771	0.991

Table 1: V_{att} on cuts scanned on different scanners

Table 2 presents the average V_{att} value measured in the experiment (E-2a), which measures V_{att} for WSI linking attacks using WSIs of consecutive slides scanned under the same conditions. Here, using the right neural network, the high feature similarity of consecutive slides also leads to successful WSI linking attacks. So, in this case as well, different WSIs can be easily linked to violate privacy of the patient.

Metric	ResNet	VGG16	Img2vec	Inception	SimCLRv2
Cosine	0.997	0.596	0.989	0.734	0.997
Euclidean	0.989	0.450	0.988	0.654	0.990

Table 2: V_{att} on consecutive slides

Finally, Table 3 presents the average V_{att} value measured in the experiment (E-2b), which measures V_{att} for WSI linking attacks using WSIs of non-consecutive slides. We have measured the effect of the spatial distance between the probe WSIs and the patient's WSI on the V_{att} value. Note that the larger distances between cuts, the fewer WSIs are available as probes; hence the table also shows the average number of probes per patient. As expected, the attack success value V_{att} decreases with the increasing physical distance of cuts.

Influence of the number of patients and probes on V_{att}. As the efficiency of our similarity-based linking method apparently depends on the number of patients and the number of probes for each patient, we have evaluated the effect of these parameters on the V_{att} value in the case of non-consecutive slides—i.e., for the experiment (E-2b).

We select a random subset of n patients, and for each patient we randomly select one slide and insert it into $\mathcal{B}$, while we insert the rest of their slides in $\mathcal{P}$; we then measure the corresponding V_{att} value. We performed this process 40 times for every number of patients n between 1 and 80. The resulting V_{att} values (median and quartiles) are presented in Figure 8a.

For the number of probes, we used only patients with 7 or more slides, leaving us with 43 patients. Given the number of probes for each patient, we have sub-sampled the number of

	Distance (mm)	Average number of probes per patient	ResNet	VGG16	Img2vec	Inception	SimCLRv2
Cosine	3	7.0	0.476	0.408	0.406	0.360	0.540
	6	5.3	0.199	0.190	0.192	0.144	0.268
	9	4.0	0.111	0.086	0.103	0.086	0.128
	12	2.7	0.064	0.048	0.065	0.051	0.078
	15	2.1	0.033	0.029	0.038	0.030	0.035
	18	1.4	0.019	0.011	0.016	0.006	0.022
Euclid	3	7.0	0.449	0.398	0.398	0.299	0.518
	6	5.3	0.195	0.168	0.188	0.132	0.246
	9	4.0	0.112	0.100	0.097	0.090	0.121
	12	2.7	0.064	0.051	0.065	0.05	0.064
	15	2.1	0.034	0.030	0.034	0.024	0.038
	18	1.4	0.015	0.016	0.013	0.008	0.021

Table 3: Influence of distance between slides on V_{att} for non-consecutive slides.

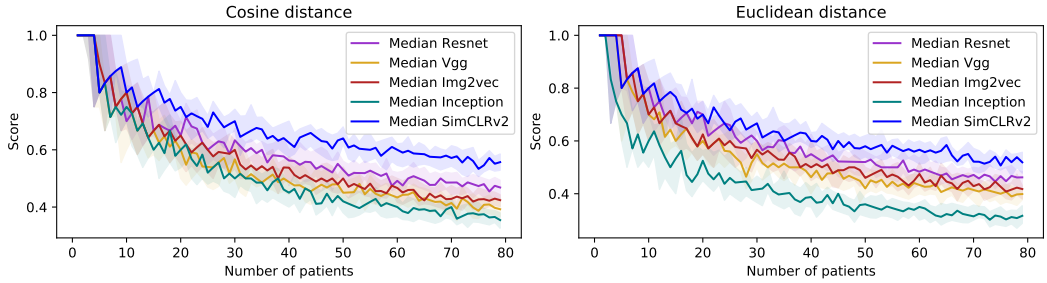
probes for each patient to the given number if the number of all probes of the patient exceeded the given number. Random slide from each sub-sample is inserted into $\mathcal{B}$ and remaining into $\mathcal{D}$. Then we have proceeded 40 times and averaged the results for each number of probes per patient between 0 and 6. Figure 8b shows the dependency of the resulting average V_{att} on the value of probes of each patient.

3.2 Results on cropped WSI linking attacks

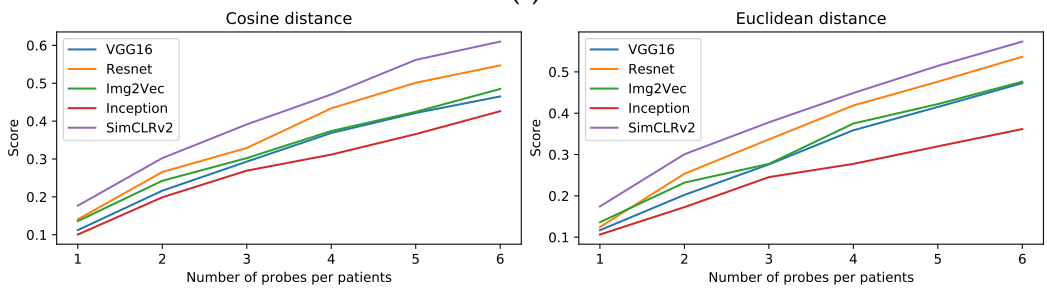
Table 4 summarizes our results on cropped WSI linking attacks measured on consecutive slides for experiment (cE-2a). V_{att} decreases only slightly when compared to results of experiment (E-2a) in Table 2 (i.e., WSIs of consecutive slides scanned under the same conditions). When comparing results of VGG16 and Inception, we can see that cropping tissue actually significantly improves V_{att}; we assume this is caused by the changes in shape/border of the tissue on consecutive slides, to which these two feature extractors seem to be more sensitive compared to ResNet and Img2vec.

Similarly, we have also tested attacks on cropped WSIs of non-consecutive slides in experiment (cE-2b); results are summarized in Table 5. Compared with results of experiment (E-2b) in Table 3 on whole non-consecutive images, there is again a slight decrease in V_{att}. Crop shifts have not been evaluated for non-consecutive slides, as that would add one more dimension to the results and this dimension was sufficiently demonstrated on consecutive slides.

We can summarize that the success of a (full) WSI linking attack using the presented methods does not depend significantly on the availability of the border and on the overall shape of the tissue, but the internal structure of the tissue is sufficient. The used attack model implementation is also relatively insensitive to small shifts in the crop area. Only after shifting by more than 50 pixels (area overlap 71% to 78% depending on the shift direction), the V_{att} value



(a)



(b)

Figure 8: Influence of (a) the number of patients and (b) number of slides (probes) on V_{att}. Median value is shown as solid line, quartiles are shown as shaded areas of corresponding color. These two experiments use non-consecutive slides from (H-2) hierarchy.

deteriorates below 0.8 using either of the similarity measures and almost any of the feature extractors.

4 Conclusions and Recommendations

WSI are a specific category of data from the privacy risk perspective. In this paper, we have primarily focused on linkage attacks, which may lead to identity disclosure or enable inferring additional information about the data subjects. As demonstrated by our experimental results achieved, with a relatively straightforward attack model, the potential to act as links across different data sets is substantial for identical WSIs, and also for WSIs generated from the same or from closely spatially related physical slides. Additional insight into risks could be obtained by organizing an international competition, where different teams could propose and compare potentially stronger attack methods. The authors propose to organize such a competition using the attack model developed in this paper as a reference and following the data release guidelines outlined later in this text.

Moreover, an aspect which has not been a specific focus of this paper, apart from generic feature extraction, is how much information can be inferred directly from WSI data—i.e., what additional information can be inferred from the "fingerprint" itself. Many AI models have been developed to infer diagnoses (survey in [17]) and disease-specific scores such as a UICC stage for colorectal cancer or a Gleason score for prostate cancer [18, 19]. But AI models have been

	Shift	ResNet	VGG16	Img2vec	Inception	SimCLRv2
Cosine	0 px	0.971	0.959	0.941	0.970	0.968
	5 px	0.971	0.958	0.941	0.969	0.979
	10 px	0.969	0.954	0.938	0.955	0.976
	15 px	0.959	0.952	0.939	0.956	0.974
	25 px	0.943	0.939	0.919	0.966	0.963
	50 px	0.903	0.905	0.898	0.909	0.924
	75 px	0.795	0.550	0.767	0.707	0.924
	100 px	0.581	0.268	0.587	0.403	0.611
Euclid	0 px	0.969	0.952	0.928	0.916	0.981
	5 px	0.968	0.955	0.927	0.914	0.985
	10 px	0.964	0.951	0.925	0.918	0.975
	15 px	0.960	0.945	0.920	0.890	0.972
	25 px	0.949	0.925	0.903	0.864	0.946
	50 px	0.937	0.897	0.855	0.811	0.915
	75 px	0.810	0.423	0.704	0.503	0.785
	100 px	0.615	0.206	0.562	0.268	0.586

Table 4: Influence of the crop area shift on V_{att} for cropped WSI of consecutive slides.

shown to predict also less apparent features, such as information on particular mutations [20–23]. Inferring such information might be used to further improve linkage attacks (e.g., if mutation information is known to the attacker) or to infer additional information about the data subjects such as clinical or genomic data. When releasing WSI data as a part of bigger data sets, the likelihood of inferring sensitive data should be taken into account, which is considered in the proposed guidelines below. This is relevant primarily when releasing the data as de facto anonymous data sets, where it can be correlated with other data sets in the future (e.g., genomic data).

To summarize, given the experimental findings and the additional considerations presented in this text, when WSI data is to be released across different public data sets—i.e., without controlled access and additional contractual responsibilities—caution needs to exercised to mitigate privacy risks. At the same time we acknowledge the need to maximize openness of data and to make large volumes of data available for the development of AI models that have the potential to significantly improve health care. Hence, we propose the risk assessment models and data release guidelines for WSIs presented in the following Section.

4.1 Proposed Risk Assessment and Data Release Guidelines for WSIs

When releasing data sets containing WSIs, it is desired to publish them *as openly as possible* to support their reuse, but also to keep the data *as closed as necessary.* In the following subsections we define the relevant terminology, discuss risk assessment aspects and propose two sets of guidelines: the first for releasing WSI as de facto anonymous data sets, and the second for

	Distance	Average number of probes per patient	ResNet	VGG16	Img2vec	Inception
Cosine	3	7.0	0.339	0.115	0.288	0.239
	6	5.3	0.154	0.071	0.149	0.115
	9	4.0	0.110	0.054	0.073	0.063
	12	2.7	0.043	0.034	0.031	0.029
	15	2.1	0.036	0.019	0.016	0.027
	18	1.4	0.010	0.01	0.002	0.013
Euclid	3	7.0	0.316	0.094	0.263	0.073
	6	5.3	0.150	0.075	0.128	0.042
	9	4.0	0.1	0.044	0.059	0.033
	12	2.7	0.051	0.029	0.022	0.02
	15	2.1	0.031	0.018	0.015	0.018
	18	1.4	0.011	0.006	0.002	0.001

Table 5: Influence of distance between slides on V_{att} for cropped WSI of non-consecutive slides.

releasing pseudonymized (or even identified) data set under a contract defining appropriate technical and organizational safeguards. Note that the privacy risk analysis in this paper is independent of any particular jurisdiction. We do, however, anchor the guidelines developed below to the terminology established by the GDPR.

Definition of *de facto anonymity*. The notion of *de facto anonymity* captures the idea that it is reasonably unlikely that the person subject of the data could be identified. We can examine the notion in more detail by reading the GDPR. In recital 26 it defines anonymous information as:

> ...information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.

To understand this definition, we need to understand what constitutes *personal data* and what makes a person *identifiable*. The GDPR defines personal data in Art. 4 (1) as:

> "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

Finally, recital 26 of the GDPR also provides important information on how to decide whether a natural person is identifiable:

> To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used… To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs and time required for identification, taking into consideration the available technology at the time of the processing and technological developments."

Therefore, the stance taken by the GDPR is clearly against the notion of anonymity as an absolute value, thus absolute anonymity once and forever with zero risk of re-identification is not required by the GDPR. Instead, certain residual risks are acceptable and can be calibrated against the sensitivity of the data in respect to the impact of privacy breaches for the data subjects. This is captured by the notion of de facto anonymity.

Further, according to the Art. 29 Working Party (predecessor of European Data Protection Board under the GDPR) even organisational measures can influence the status of anonymity,[4] since identifiability depends on the background knowledge of potential attackers: the same data might be anonymous in one setting and personal data in another. Examples of such organizational measures include: access control and contractual obligation to not re-identify research participants, to not share the data with third parties, and to make them internally accessible only under confidentiality obligations, provided that the contractual party is reliable and able to fulfill those obligations.

Guidelines for releasing WSI as de facto anonymous data. The following technical and organizational measures shall be considered before releasing WSI data in de facto anonymous data sets:

- Each WSI shall be stripped of any patient-related metadata. Technical metadata related to scanning processes should be reviewed and minimized for the given purpose (e.g., removing location identifiers if present, but retaining information on scanning parameters).
- If a WSI is released into more than one de facto anonymous data set, corresponding records (i.e., WSI and data linked to it) in these data sets shall be considered linked.
- If the same WSI is released in a de facto anonymous data set and other non-open datasets, gaining information from the de facto anonymous data set is technically trivial for the recipients of the non-open data set and must be considered from the risk assessment perspective.
- WSIs to be released in a de facto anonymous data set and the originating slide must be dissociated from the patient records. How the dissociation is done depends on the context of the de facto anonymization. It can be done using a legally enforceable contract permanently preventing the recipient of the data from accessing the link. If this is not possible, the dissociation has to be done by removing all references between the WSIs/originating slides and any patient records in the data holder's/controller's information systems.

[4] Art. 29 WP opinion 136, concept of personal data, p. 17

- If the WSI is a part of a series of consecutive slides, the same rules on dissociation apply to the whole consecutive WSIs stack and corresponding original slides.
- The probability of successfully inferring information from the WSIs shall be considered and in case that such information could practically lead to singling out a patient, the WSIs shall not be released as anonymous material. Examples include rare cancer diagnoses and rare mutations. Qualified risk assessments need to be carried out according to state-of-the-art methodologies.

When releasing data as de facto anonymous, a legally enforceable contract shall be in place between the data controller and the data recipient, which prohibits re-identifying data subjects and/or inferring data about any specific individuals. This is a legal measure to mitigate residual risks, such as risks arising from the development of novel re-identification techniques for WSI data. Such legal measures may include limiting the purpose of the data processing as well as setting time constraints for the processing in order to manage risks related to the development of new methods in the future.

When releasing WSI data in an open data set (e.g. publicly downloadable on the internet), at least a lightweight contract, such as a licence agreement for the dataset, shall be implemented to which the recipient of the dataset must actively agree.

Guidelines for releasing WSIs as personal data. When there is a legal basis for processing personal data for a particular purpose, it is recommended to release the WSIs and any other necessary linked personal data using a legally enforceable contractual basis. Under the GDPR, this can be a data transfer agreement (i.e., establishing controller to controller data transfer) or processing on behalf of the data controller agreement. The data set is typically pseudonymized, i.e., direct identifiers are replaced by a pseudonym for each patient, and minimized for the given purpose. The contracts need to restrict the purpose of use and define requirements data processing.

This is also advantageous because the research results can then be fed back to the donors, e.g. in the form of acting on incidental findings to prevent the development of a disease not yet known to the donor.

Information leakage from AI models. When AI models are developed using WSI data, it also needs to be ensured that the model does not leak personal information learned from training data (see, for instance, the work by Shokri et al. [24]). Methods like PATE [25] can be used to mitigate these risks.

Provenance information management and release. Provenance information is an important aspect of trustworthy data with defined and analyzable quality [26]. For personal/pseudonymous data sets, provenance information can lead back to the originating biological material, but the link to the donating patient can typically only be resolved by authorized personnel at the source organization. When releasing de facto anonymous data sets, provenance information can only start with the anonymization process and cannot link to the originating data (this link must be intentionally and permanently destroyed).

Acknowledgements

This work has been supported by EOSC-Life project supported by EU Horizon 2020, grant agreement no. 824087, as a part of WP1 Demonstrators under APPID 1228 "Cloudification of BBMRI-ERIC CRC-Cohort and its Digital Pathology Imaging"; by the Czech Ministry of Health (MMCI 00209805) and the Czech Ministry of Education, Youth and Sports (LM2018125 – BBMRI-CZ); the Austrian Federal Ministry for Education, Science and Research (BMBWF-10.470/0010-V3c/2018; BBMRI.at); the SVDC project (funded by the Sardinian Regional Authority). Computational resources were supplied by the project 'e-Infrastruktura CZ' (e-INFRA LM2018140) provided within the Projects of Large Research, Development and Innovations Infrastructures program.

Conflict of interest

The authors declare no conflict of interests.

Ethics check

The research has been under ethics vote at Masaryk Memorial Cancer Institute, no. 2021/3375/MOU (MOU 385 920), and at Medical University Graz, no. 1072/2019. This work has not attempted to identify nor identified any data subjects, it solely performed the linking-based risk analysis solely on WSI data.

References

1. Zelic, R. *et al.* Interchangeability of light and virtual microscopy for histopathological evaluation of prostate cancer. en. *Scientific Reports* **11,** 3257. ISSN: 2045-2322 (Dec. 2021).
2. Holzinger, A. *et al. Machine Learning and Knowledge Extraction in Digital Pathology Needs an Integrative Approach* in *BIRS-IMLKE* (2015).
3. *Artificial Intelligence and Machine Learning for Digital Pathology* (eds Holzinger, A., Goebel, R., Mengel, M. & Müller, H.) (Springer International Publishing, 2020).
4. Ehteshami Bejnordi, B. *et al.* Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA* **318,** 2199–2210. ISSN: 1538-3598 (Dec. 2017).
5. Bandi, P. *et al.* From Detection of Individual Metastases to Classification of Lymph Node Status at the Patient Level: The CAMELYON17 Challenge. *IEEE transactions on medical imaging* **38,** 550–560. ISSN: 1558-254X (2019).
6. Gutman, D. A. *et al.* Cancer Digital Slide Archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *J Am Med Inform Assoc* **20,** 1091–1098 (2013).

7. Malin, B., Loukides, G., Benitez, K. & Clayton, E. W. Identifiability in biobanks: models, measures, and mitigation strategies. *Hum Genet* **130,** 383–392 (Sept. 2011).

8. Feit, J., Kempf, W., Jedlicková, H. & Burg, G. Hypertext atlas of dermatopathology with expert system for epithelial tumors of the skin. *J Cutan Pathol* **32,** 433–437 (July 2005).

9. Jezová, M., Múcková, K., Soucek, O., Feit, J. & Vlasín, P. Hypertext atlas of fetal and neonatal pathology. *Diagn Pathol* **3 Suppl 1,** S9 (July 2008).

10. Lianas, L. *et al.* Cy-TEST - A new platform for training and testing in cytopathology. *Diagnostic Pathology* **2.** ISSN: 2364-4893 (2016).

11. Veeling, B. S., Linmans, J., Winkens, J., Cohen, T. & Welling, M. Rotation Equivariant CNNs for Digital Pathology. arXiv: `1806.03962 [cs.CV]` (June 2018).

12. Deng, J. *et al. Imagenet: A large-scale hierarchical image database* in *2009 IEEE conference on computer vision and pattern recognition* (2009), 248–255.

13. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **45** (IEEE, June 2016), 770–778. ISBN: 978-1-4673-8851-1.

14. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings,* 1–14 (2015).

15. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. arXiv: `1512.03385 [cs.CV]` (2015).

16. Chen, T., Kornblith, S., Swersky, K., Norouzi, M. & Hinton, G. Big Self-Supervised Models are Strong Semi-Supervised Learners. *arXiv preprint arXiv:2006.10029* (2020).

17. Pocevičiūtė, M., Eilertsen, G. & Lundström, C. in *Artificial Intelligence and Machine Learning for Digital Pathology* (eds Holzinger, A., Goebel, R., Mengel, M. & Müller, H.) 56–88 (2020).

18. Wulczyn, E. *et al.* Interpretable survival prediction for colorectal cancer using deep learning. *npj Digital Medicine* **4** (Apr. 2021).

19. Wulczyn, E. *et al.* Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading. *Communications Medicine* **1** (June 2021).

20. Schaumberg, A., Rubin, M. & Fuchs, T. H&E-stained Whole Slide Image Deep Learning Predicts SPOP Mutation State in Prostate Cancer. *bioRxiv* **XXX,** 1–14 (2018).

21. Chang, P. *et al.* Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *American Journal of Neuroradiology* **39,** 1201–1207. ISSN: 1936959X (2018).

22. Xu, H., Park, S., Lee, S. H. & Hwang, T. H. Using transfer learning on whole slide images to predict tumor mutational burden in bladder cancer patients. *bioRxiv,* 1–17. ISSN: 2692-8205 (2019).

23. Schmauch, B. *et al.* A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nature Communications* **11,** 1–15. ISSN: 20411723 (2020).

24. Shokri, R., Stronati, M., Song, C. & Shmatikov, V. *Membership inference attacks against machine learning models* in *2017 IEEE symposium on security and privacy (SP)* (2017), 3–18.

25. Papernot, N. *et al. Scalable Private Learning with PATE* 2018. arXiv: `1802.08908 [stat.ML]`.

26. Holub, P. *et al.* Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health. *Biopreserv Biobank* **16,** 97–105 (Apr. 2018).