ORIGINAL PAPER

Modeling and simulation of diffusion and reaction processes during the staining of tissue sections on slides

Johannes D. M. Menning[1] · Thomas Wallmersperger[1,3] · Matthias Meinhardt[2] · Adrian Ehrenhofer[3,1]

Abstract

Histological slides are an important tool in the diagnosis of tumors as well as of other diseases that affect cell shapes and distributions. Until now, the research concerning an optimal staining time has been mainly done empirically. In experimental investigations, it is often not possible to stain an already-stained slide with another stain to receive further information. To overcome these challenges, in the present paper a continuum-based model was developed for conducting a virtual (re-) staining of a scanned histological slide. This model is capable of simulating the staining of cell nuclei with the dye hematoxylin (C.I. 75,290). The transport and binding of the dye are modeled (i) along with the resulting RGB intensities (ii). For (i), a coupled diffusion–reaction equation is used and for (ii) Beer–Lambert's law. For the spatial discretization an approach based on the finite element method (FEM) is used and for the time discretization a finite difference method (FDM). For the validation of the proposed model, frozen sections from human liver biopsies stained with hemalum were used. The staining times were varied so that the development of the staining intensity could be observed over time. The results show that the model is capable of predicting the staining process. The model can therefore be used to perform a virtual (re-)staining of a histological sample. This allows a change of the staining parameters without the need of acquiring an additional sample. The virtual standardization of the staining is the first step towards universal cross-site comparability of histological slides.

Introduction

Cancer is one of the leading causes of death worldwide today (Ferlay et al. 2020). In order to discover and treat cancer appropriately, it is necessary to examine the histological characteristics of tumor cells (Jones et al. 2015). The two main ways to diagnose the presence of cancer and its characteristics, such as the degree of malignancy are (i) the staining of histological sections (Falk et al. 2018; Giuliano et al. 2011; Hiddemann and Bartram 2010; Veuthey et al. 2014) and (ii) radiological techniques, e.g., magnetic resonance imaging (Dong et al. 2017). It is anticipated that pathologists will receive support in the diagnosis of diseases from models based on artificial intelligence methods (Ari and Hanbay 2018; Dong et al. 2017; Kather et al. 2019; Sharma et al. 2014). Additionally, methods without the need of preparing classical histological slides were developed (Hollon et al. 2020).

The goal of staining is to create a visual distinction between the components of the tissue (see Fig. 1). To obtain a general overview, the cell nuclei and cytoplasm are stained with an appropriate dye.

The most common tissue stains for that purpose are hematoxylin (stains mainly the nuclei) and eosin (stains

Matthias Meinhardt and Adrian Ehrenhofer have contributed equally to this work.

✉ Matthias Meinhardt
matthias.meinhardt@uniklinikum-dresden.de

✉ Adrian Ehrenhofer
adrian.ehrenhofer@tu-dresden.de

1 Technische Universität Dresden, Institute of Solid Mechanics, George-Bähr-Straße 3c, 01069 Dresden, Germany

2 University Hospital Carl Gustav Carus Dresden, Institute of Pathology, Fetscherstraße 74, 01307 Dresden, Germany

3 Technische Universität Dresden, Dresden Center for Intelligent Materials, School of Engineering Sciences, George-Bähr-Straße 3c, 01069 Dresden, Germany
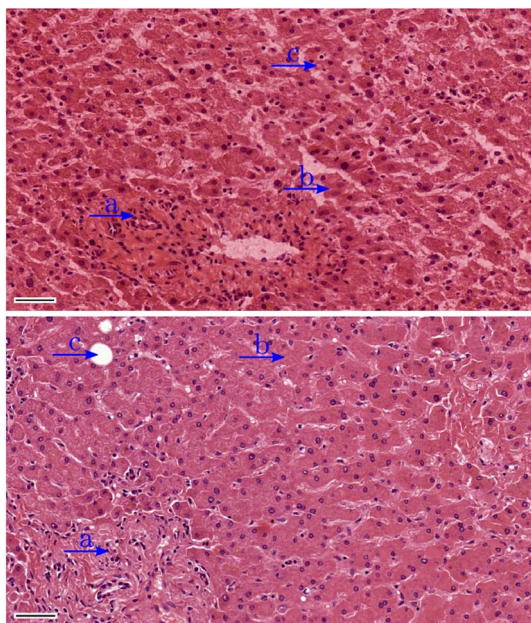
Fig. 1 Comparison of a frozen section (*top*) with a paraffin section (*bottom*) from the same patient sample. The region denoted by **a** is a portal field with bile ducts and lymphocytes, **b** denotes liver cells which are arranged in cell strands (trabeculae) and **c** is a fat vacuole. The *scale bar* is 50 μm

mainly the cytoplasm) (Lang 2006; Mulisch and Welsch 2010; Suvarna et al. 2019; Veuthey et al. 2014).

The aim of the present paper is to model and simulate the nuclei staining of histological sections with the dye hematoxylin. A histological slide, which was already stained with dye A, could therefore be virtually re-stained with a different dye B. Such a model is feasible because each cell in an organism can be characterized by the structure of its nuclei and cytoplasm (Mulisch and Welsch 2010). Additionally, the dye reacts always in the same way with the respective tissue (Lang 2006).

Modeling and simulation of dyeing in general has been studied mainly with respect to textile fiber dyeing (de Souza et al. 2007; Lin 1992; Reddy et al. 1995). Reddy et al. (1995) used the Langmuir model and compared that to a diffusion equation. de Souza et al. (2007) and Lin (1992) used coupled diffusion–reaction equations. There are also examples like by Winzek and Baumgärtel (1988) who simulated the staining of histological slides or by Irion et al. (1993) who modeled the staining of single-cell organelles.

In the current work, the mechanics of the staining of histological sections is investigated via the methods of building a simplified model and conducting numerical simulations. After a short introduction into the basics of histological staining, the model to describe the diffusion and reaction processes during the staining of histological sections is presented. This model allows the determination of the resulting

stain for each nucleus individually, since it is important for an investigator to look at the entire tissue section and not just individual nuclei. Therefore, a workflow was created to derive a two-dimensional representation of the geometry of the scanned histological slide. This geometry is then used to create a finite element mesh for the sample. This procedure is described in the following section, together with the discretization of the problem and the applied initial conditions and boundary conditions. The proposed model is then compared to real histological slides that were created at the Institute of Pathology at the University Hospital Carl Gustav Carus Dresden.

Materials and methods

Theoretical framework and continuum-based model

Here, we give a brief introduction to the dye used for the experiments and simulations and how it binds to the tissue sections. Afterwards, the proposed model for the computation of the dye concentration and the resulting color is described. The model is based on continuum mechanics, i.e., the local concentration of the dye changes due to transport that is caused by diffusion.

Hematoxylin

Hematoxylin is a commonly used dye to stain nuclei (Prentø 2001) and is often used in a combination with eosin (Mulisch and Welsch 2010; Suvarna et al. 2019) but also as a counter stain for immunohistochemical staining (de Cea and Nie 2018; Suvarna et al. 2019). For this reason, the staining of a histological section with hematoxylin was modeled.

The actual dye is not hematoxylin but a combination of the oxidized hematoxylin (called hematein) and a mordant with additional components. There are diverse recipes with varying amounts of hematein, mordant, and other components (Prentø 2001, 2009). In our investigation, Mayer's hemalum was used for the experimental tissue sections as well as for the numerical simulation.

Physico-chemical binding theory of tissue staining

Several theories exist that describe the processes involved in tissue staining. The most commonly used theory is the physical–chemical bonding theory, which is based on electrostatic attraction (Mulisch and Welsch 2010; Veuthey et al. 2014). It is assumed that the dye and the tissue have an acidic or basic pH value, which results in a negative or positive charge. After the dye has interacted with the tissue due to the electrostatic attraction, a chemical bond is formed (Mulisch and Welsch 2010; Prentø 2001, 2009; Veuthey

et al. 2014). Please note that a dye must have complementary properties to the tissue in order to be able to form a chemical bond. This is necessary in order to compete against other molecules (Prentø 2001).

Histological changes of tumors

For tumor cells, a de-differentiation occurs, which is accompanied by changes in shape of the nucleus that becomes often enlarged and irregularly shaped and is hyperchromatic. This is also referred to as pleomorphism. In addition, the cytoplasm of the tumor cell may exhibit a staining different from normal cells.

Transport and binding of the dye

In literature concerning the mechanics of staining, the diffusion is described to be the dominant factor for the transport of the dye (Goldstein 1980; Irion et al. 1993; Siedel and Zimmermann 1995; Winzek et al. 1987). Therefore, a diffusion term is used to model the transport of the dye. Inside the cell, no difference regarding the diffusion coefficient is made between its single components, e.g., the membranes, cytoplasm, or nucleus. As a consequence, the diffusion in the cell can be modeled as a single step. For a better understanding of the staining process, a dye basin with a histological slide as well as a simplified model of a cell is depicted in Fig. 2. Usually, the glass slides with the attached tissue sections are moved inside the dye basin. Thus, the dye solution is always in motion and is getting mixed. This leads to a constant dye concentration at the surface of the histological slide. Therefore, diffusion is modeled only in the tissue and not also inside the dye basin.

Once the dye has reached its binding sites due to the electrostatic attraction, it forms a chemical bond with the

respective tissue (see "Physico-chemical binding theory of tissue staining").

The binding of the dye to the tissue is a reversible reaction (Ferus-Comelo 2002; Winzek and Baumgärtel 1988). Therefore, the binding of the dye can be represented as a pseudochemical reaction

$$A + B \leftrightarrow C \tag{1}$$

where A is the free, unbound dye concentration, B the tissue with which the dye can form a bond, and C is the dye which is bound to the tissue.

The staining can be considered to be a two-step process (Prentø 2009). The first step is the diffusion of the dye into the tissue section and the second step is the binding of the dye to the tissue components. For the simulation of both steps, a diffusion–reaction equation and a reaction equation are fully coupled to form the resulting system of equations. The detailed derivation can be found in Menning (2021).

$$\frac{\partial c_f}{\partial t} = D\nabla^2 c_f - k_{on} c_f \left(c_b^{max} - c_b \right) + k_{off} c_b \tag{2}$$

$$\frac{\partial c_b}{\partial t} = k_{on} c_f \left(c_b^{max} - c_b \right) - k_{off} c_b \tag{3}$$

The used material parameters are the diffusion coefficient D, the rate constant which determines the binding rate k_{on}, the rate constant which determines the dye release rate k_{off}, and the maximal concentration of dye c_b^{max} that the respective part of the cell, in this case the nucleus, can bind. The general variables in Eq. (1) can now be defined with the used variables

$$c_f + \left(c_b^{max} - c_b \right) \rightleftharpoons c_b \tag{4}$$

Please note that Eq. (4) is not an exact chemical formula; it is only for better representation. The unknown functions in Eqs. (2) and (3) are the concentration of the unbound dye c_f (which is free to move) and the concentration of the bound dye c_b. The latter describes the concentration of the dye which has formed a chemical bond with the tissue. Therefore, c_b is no longer able to diffuse and thus no diffusion term is needed in Eq. (3). Additionally, to the diffusion in the first step (Prentø 2009) describes an ion exchange. This ion exchange is not explicitly modeled. It determines how and where the dye is able to bind to the tissue. Through the common use of Mayer's hemalum, it is already known that the dye can bind to almost all nuclear acids (Prentø 2001). Equations (2) and (3) are used to model the transport and the binding of a dye to a histological section. Equations (2) and (3) are similar to the equations used by Lin (1992) to model the staining of textile fibers.
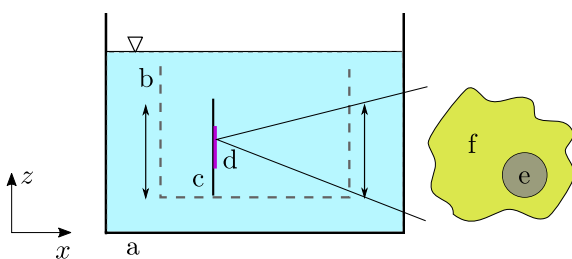

Fig. 2 Illustration of a dye basin (**a**) in which the slide rack (**b**) is submerged. A histological slide (**c**) with a tissue sample (**d**) represents multiple histological sections situated on the slide rack. A zoom of this tissue sample shows a simplified model of a cell used in the current work. In this simplification, the cell consists of only two components, the cell nucleus (**e**) and the cytoplasm (**f**)

RGB color model

There exist different models to mathematically describe a color (Bejnordi et al. 2015; Ruifrok et al. 2001; Van der Laak et al. 2000). The model used in this paper is the RGB model, which is an additive color model (Heid and Reith 2010). The three intensities red (R), green (G), and blue (B) span a vector space whose axes are the three intensities (Fig. 3). Typically, the values range between 0 to 1 or from 0 to 255. The latter range was used in the current work. A combination of (0,0,0) represents the color black and if each intensity is maximal, i.e., 255, the color white is represented (Van der Laak et al. 2000). That is, the lower the intensities, the darker the color.

Derivation of a color from the dye concentration

Using the equations given in "Transport and binding of the dye", it is possible to calculate the concentration of the dye bound to the tissue c_b. To derive the resulting color from this concentration, Beer–Lambert's law.

$$I = I_0 \exp(-\varepsilon c_b h) \tag{5}$$

is used. It links the intensity of the light after passing the slide I with the concentration of the bound dye c_b via the extinction coefficient ε, the thickness of the slide h and the intensity of the light before it enters the specimen I_0 (Bettinger and Zimmermann 1991; Gavrilovic et al. 2013; Ruifrok et al. 2001). The extinction coefficient ε is a material parameter of the dye. It determines which wavelengths λ of the light are absorbed by the dye and are thus no longer present in the light leaving the slide. Please note that the Beer–Lambert's law is only valid for dyes that absorb the light like hematoxylin or eosin. Beer–Lambert's law cannot be applied for dyes that reflect light (Gavrilovic et al. 2013).

The intensities I and I_0 as well as the extinction coefficient ε are dependent on the wavelength λ. The thickness
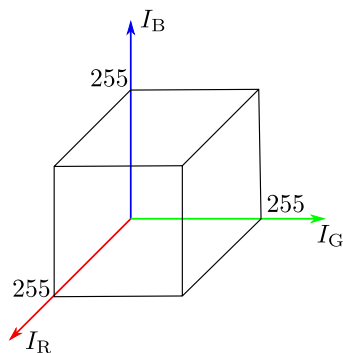

Fig. 3 Representation of the color space defined by the red, green, and blue intensity according to Van der Laak et al. (2000)

h of the slide is a parameter of the slide's production process. Following Van der Laak et al. (2000), Eq. (5) can be formulated individually for the three colors red (R), green (G), and blue (B)

$$I_R = I_{0,R} \exp(-\varepsilon_R c_b h) \tag{6}$$

$$I_G = I_{0,G} \exp(-\varepsilon_G c_b h) \tag{7}$$

$$I_B = I_{0,B} \exp(-\varepsilon_B c_b h) \tag{8}$$

Equations (6)–(8) can be used to calculate the intensities I_R, I_G and I_B and required to derive a resulting color. The extinction coefficient ε can be (i) determined independently by experimental investigations, (ii) taken from the literature or be (iii) calculated from data of stained slides, as has been done in the present work.

Numerical implementation

For the implementation of the introduced equations, the open-source software FEniCS (Alnæs et al. 2015) was used. After the discretization, the initial conditions and the boundary conditions are given in "Initial conditions and boundary conditions". After that, the used material parameters, are determined in "Material parameters". In "Finite element mesh from a scanned slide", the workflow for deriving a FE mesh based on a scanned slide is described.

Discretization

The finite element method is used to discretize and solve Eqs. (2) and (3). The required weak formulations are

$$\int_\Omega \dot{c}_f \delta c_f \mathrm{dV} = \int_{\partial\Omega} n \cdot (D\nabla c_f \delta c_f)\mathrm{dA} - \int_\Omega D\nabla c_f \cdot \\ \nabla \delta c_f \mathrm{dV} - \int_\Omega k_{on} c_f(c_b^{max} - c_b)\delta c_f \mathrm{dV} + \int_\Omega k_{off} c_b \delta c_f \mathrm{dV} \tag{9}$$

$$\int_\Omega \dot{c}_b \delta c_b \mathrm{dV} = \int_\Omega k_{on} c_f(c_b^{max} - c_b)\delta c_b \mathrm{dV} - \int_\Omega k_{off} c_b \delta c_b \mathrm{dV} \tag{10}$$

with the two weighting functions δc_f and δc_b and the integrals over the area A and the volume V. **n** represents the normal vector at the boundary in outwards direction. Typical dimensions of a histological section are 20 mm × 15 mm in the plane (Mulisch and Welsch 2010) and only a few micrometer in thickness. For the sections actually used in the experiments, a cutting thickness of $h = 6\mu m$ was set at the microtome. Due to these striking differences in dimensions, diffusion over h, i.e., along the z-axis (Fig. 4 (top)) will be much faster than in the xy-plane. Therefore, the assumption was made that diffusion in the xy-plane can be neglected.
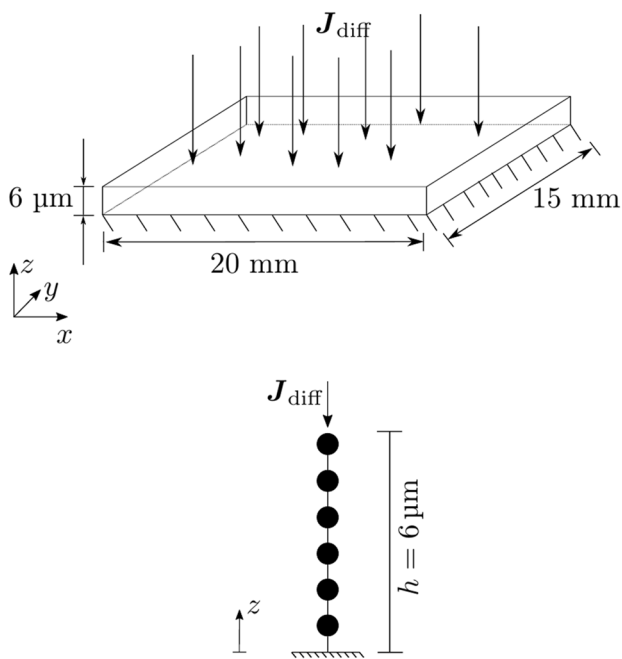
Fig. 4 *Top panel*: Schematic representation of a histological section with the diffusional flux J_{diff}. At the bottom, the tissue is situated on a glass slide which blocks the transport of the dye, represented by the *hatching*. *Bottom panel*: Representation of the discrete one-dimensional FE mesh used to calculate dye transport

Thus, the problem is considered to be *one-dimensional* in z-direction. In Fig. 4 (bottom), a schematic representation of the used one-dimensional finite element mesh (FE mesh) is shown.

For the time discretization, the backward Euler finite difference method is used

$$\dot{c}_{\text{f}} = \dot{c}_{\text{f}}^{n+1} \approx \frac{c_{\text{f}}^{n+1} - c_{\text{f}}^{n}}{\Delta t} \tag{11}$$

$$\dot{c}_{\text{b}} = \dot{c}_{\text{b}}^{n+1} \approx \frac{c_{\text{b}}^{n+1} - c_{\text{b}}^{n}}{\Delta t} \tag{12}$$

where n is the actual point in time and t is the time. The used time step size is $t_{n+1} - t_n = \Delta t = 0.5$s.

Initial conditions and boundary conditions

The simulation starts right before the slides are placed into the dye basin. Hence, the concentration of the free and of the bound dye are both equal to zero at $t = 0$s, i.e., $c_{\text{f}}(z, t = 0\text{s}) = 0\,\text{mol}\,\text{m}^{-3}$ and $c_{\text{b}}(z, t = 0\text{s}) = 0\,\text{mol}\,\text{m}^{-3}$.

At its surface, where the tissue section is in contact with the dye solution, a constant concentration of the free, unbound dye is assumed, $c_{\text{f}}(z = 6\,\mu\text{m}, t) = \overline{c}_f$ with $\overline{c}_f$ being

the concentration of the dye used in the respective recipe. At the tissue section base, due to the restricting glass slide, the flux is set to zero, i.e., $J_{\text{diff}}(z = 0\,\text{m}, t) = 0\,\text{s}^{-1}\,\text{m}^{-2}$. In Fig. 4, these two conditions are visualized.

For the bound dye concentration c_{b}, no boundary condition is necessary.

Material parameters

The material parameters used for the comparison and the validation of the model proposed with the experimental data were calibrated from two of five different staining times, with the exceptions of $I_{0,i}$ with $i = R, G, B$. With the assumption of white light passing through the slide, $I_{0,i}$ with $i = R, G, B$ can be set to $I_{0,i} = 255$.

A combination of Merck's Mayer's hemalum and SAV's Mayer's hemalum with a mixing ratio of 1:2 in favor of SAV's hemalum recipe was used to stain the tissue sections attached to the slides. The hematin and the mordant form a bond in a ratio of 1:1 (Bettinger and Zimmermann 1991). Therefore, the concentration of the hematin and the same amount of the mordant were summed up. This is done for the recipe of Merck and SAV. From the calculated values, the weighted average is taken. The result is referred to as $\overline{c}_f$.

Because no rate constants were found in the literature concerning the binding and the release of hemalum to and from the tissue, two assumptions were made. The first assumption is that the diffusion is the time limiting process (Goldstein 1980; Winzek and Baumgärtel 1988) and the second one that the dye release will be much slower than the binding (Siedel and Zimmermann 1995). For the determination of the rate constants, the inequation

$$t_{\text{binding}} \ll t_{\text{diffusion}} \ll t_{\text{dyerelease}} \tag{13}$$

following Fibich et al. (2005) was set up. For the binding time, the diffusion time and the dye release time the following equations were used.

$$t_{\text{binding}} = \left(k_{\text{on}} c_{\text{b}}^{\text{max}}\right)^{-1}; \quad t_{\text{diffusion}} = \frac{h^2}{2D}$$

$$t_{\text{dyerelease}} = k_{\text{off}}^{-1}$$

The combination of these equations and the inequation (13) leads to the used inequation

$$\frac{k_{\text{off}}}{c_{\text{b}}^{\text{max}}} \ll \frac{2D}{h^2 c_{\text{b}}^{\text{max}}} \ll k_{\text{on}} \tag{14}$$

for the determination of the rate constants. In order to achieve that k_{on} is much larger and k_{off} much smaller than $\frac{2D}{h^2 c_{\text{b}}^{\text{max}}}$ the resulting values are to be multiplied by a large

value. In the present work, $\frac{2D}{h^2 c_\mathrm{b}^{\max}}$ was multiplied by 500 in order to be physically consistent.

Note that the rate constants are no longer variable parameters but depend on the diffusion coefficient D. The parameter k_on also depends on the maximum number of binding sites $c_\mathrm{b}^{\max}$.

The result of staining the nuclei in a histologic section can vary with some nuclei being more intensely stained than others. A possible explanation for this could be the difference in volume of the nuclei. If the number of binding sites per nucleus is the same, a change in volume results in a change in concentration. Another reason which is especially important when dealing with tumors is the fact that tumor cells in comparison to normal parental cells can have a different number of chromosomes or chromosome parts (Haroske et al. 2001). For the staining with hemalum, more chromosomes or more chromosome parts are equivalent to more binding sites and therefore a more intense staining is possible, depending on the used staining time. The amount of dye which each nucleus can bind determines how intense the virtual staining will be. To be able to consider this effect, each nucleus is assigned an individual maximal concentration of free binding sites $c_\mathrm{b}^{\max}$ for the dye. Therefore, an approximation is needed for $c_\mathrm{b}^{\max}$. Thus, the assumption was made that the aluminum ions of the hemalum will bind to the phosphor atoms of the cell nuclei (Lang 2006; Prentø 2001; Veuthey et al. 2014). The number of phosphor atoms were then estimated with the following consideration. In a nucleus are two sets of chromosomes. Each chromosome set is built up of 3.2×10^9 base pairs. A base pair consists of two nucleotides, which in turn consist, among other things, of a phosphate residue (Rassow et al. 2012). This results in 1.28 $\times 10^{10}$ phosphor atoms in the DNA per nucleus and therefore 1.28×10^{10} free binding sites for the hemalum per nucleus. In order to obtain a concentration, the number of particles must be related to a volume. In this paper, the assumption

was made that the nucleus has the shape of a sphere and that due to the high aluminum concentration used in the recipe for the hemalum, the dye molecules can bind to all binding sites. For the diameter of the sphere, $d = 6\,\mu\mathrm{m}$ is chosen, which leads to a concentration of $c_\mathrm{b}^{\max} = 187$ mol m^{-3}. Please note that the diameters can also be smaller than the thickness of the slide. However, to get a fix point, it was assumed that most of the nuclei have a diameter that is approximately the thickness of the section (Winzek and Baumgärtel 1988). Lang (2006) reported that the aluminum ions can bind to either one or two phosphor atoms. To take this into account, $c_\mathrm{b}^{\max}$ was divided by the mean value, i.e., 1.5. Since a number of assumptions were made in determining $c_\mathrm{b}^{\max}$ such as the shape and size of the nuclei or the number of binding sites, a range of values were used for the test calculations (Table 1). Due to the dependency of k_on on $c_\mathrm{b}^{\max}$, k_on is also specified through a value range. The diffusion coefficient was fitted using the mean RGB intensities of the slide from the experiments, which was stained for $t = 25\,\mathrm{s}$. Please note that the determination of the parameters can also be done at a different time point. However, our preliminary studies have shown that the best results can be obtained at an early time point, i.e., when the diffusion process is not yet completed.

The extinction coefficients were fitted using the mean RGB intensities from the experiments for the slide, which was stained for $t = 100$ s. At this time point, the staining is completed (Fig. 6) and therefore, the assumption was made that the concentration of bound dye c_b is equivalent to the maximum concentration of free binding sites $c_\mathrm{b}^{\max}$. Following that, the Beer–Lambert's law was converted to the extinction coefficients ε_i for $i = R,G,B$. With the analytically determined $c_\mathrm{b}^{\max}$, the thickness of the section h and the mean RGB intensities, all unknowns are determined. Of note, the thereby determined values of ε are smaller than the values for ε according to Bettinger and Zimmermann (1991). They specify a maximal molar extinction coefficient as well as a curve for the extinction over the wavelength for multiple

Table 1 Material parameters used in the simulation

Parameter	Description	Value	Unit
Binding rate constant	k_on	$0.34\ldots2.52$	m^{3}mol^{-1}s^{-1}
Dye release constant	k_off	0.013	s^{-1}
Diffusion coefficient	D	$4.53 \cdot 10^{-12}$	m^{2}s^{-1}
Maximal dye concentration that the tissue can bind	$c_\mathrm{b}^{\max}$	$50\ldots376$	molm^{-3}
Dye concentration in the dye basin	$\bar{c}_\mathrm{f}$	10.67	molm^{-3}
Extinction coefficient hemalum red	ε_R	1585	m^{2}mol^{-1}
Extinction coefficient hemalum green	ε_G	2996	m^{2}mol^{-1}
Extinction coefficient hemalum blue	ε_B	1111	m^{2}mol^{-1}
Intensity of the incoming light red	$I_{0,\mathrm{R}}$	255	Wm^{-2}
Intensity of the incoming light green	$I_{0,\mathrm{G}}$	255	Wm^{-2}
Intensity of the incoming light blue	$I_{0,\mathrm{B}}$	255	Wm^{-2}

hemalum concentrations with different preparation steps. One explanation for the differences could lie in the fact that the maximal molar extinction coefficient was measured in a solution and not in a tissue section as it was done in the current work.

For the calibration of the material parameters, the data from two staining time points were used. At $t = 25$s, the equilibrium was not yet reached and the dynamic parameters could be fitted. In the current work, these parameters are the diffusion coefficient D and also to a certain extend the rate constants k_{on} and k_{off} via inequation (14). However, this procedure is only needed because we are not aware of any reliable sources for these parameters. The same applies for the extinction coefficients. If these parameters are all determined through independent experiments, the remaining material parameter that has to be determined is c_b^{max}. It is important to perform this with a slide where the equilibrium (end state) of the staining is achieved. If the parameter is fitted at e.g., $t = 25$ s the resulting c_b^{max} would not be the actual maximal concentration of binding sites. This would

only make sense if the staining would normally be ceased at this time point because the desired staining has reached saturation. The physical meaning of c_b^{max} would then change from the (total) maximal concentration of binding sites to the maximal concentration of binding sites needed to achieve the desired staining.

The maximal concentration of free binding sites c_b^{max} together with the extinction coefficients ε_i for $i = R, G, B$ are the parameters, which determine the resulting stain. While the extinction coefficients are the same for the whole slide, c_b^{max} can vary for the nuclei, depending on their size. The material parameters used for the numerical simulation are summarized in Table 1.

Finite element mesh from a scanned slide

The simulations are conducted on a one-dimensional mesh (Fig. 4). In Fig. 5, the workflow for the creation of a FE mesh based on a real slide is presented. This mesh is used only to record the projected simulation results to create a
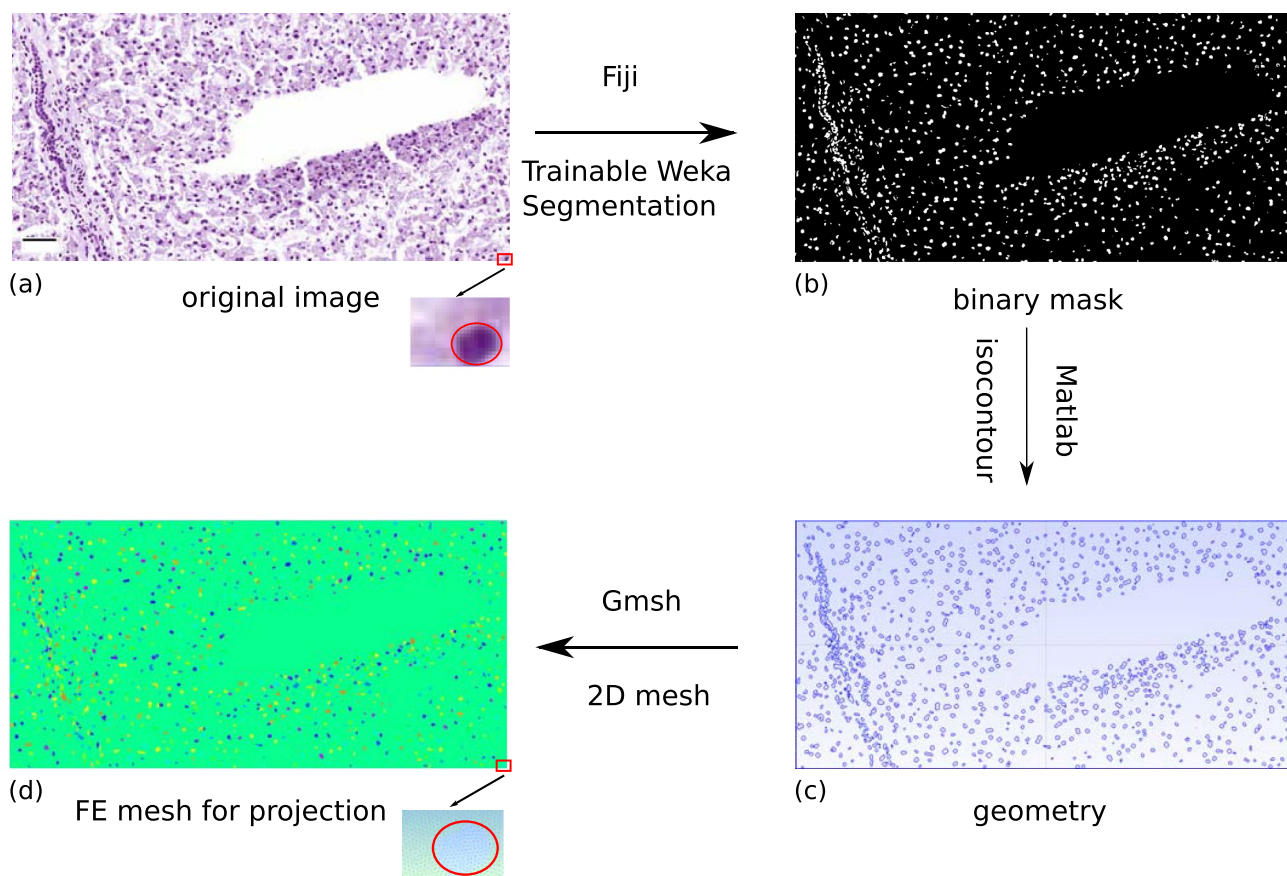

Fig. 5 Presentation of the important steps during the creation of an FE mesh based on a section of a scanned histological slide. The slide from which a section is shown in **a** was stained for 13 s with Mayer's hemalum. In **b**, the binary mask for the cell nuclei is depicted. The geometry is shown in **c**. In **d**, the FE mesh for the projection is pre-

sented. Additionally, a single nucleus (*red outlined*) is shown at high magnification. The resulting mesh for this nucleus is shown at the bottom of the image of the "FE mesh for projection". The *scale bar* is 50 μm

two-dimensional image (Supplementary Fig. 5). Thus, no computations are carried out with this mesh. In Fig. 5a, a single nucleus is depicted at high magnification. Zooming in on a section of the image shows the nucleus approximated by finite elements (Fig. 5d).

The process starts with a scan of the histological section (Fig. 5a) for which the virtual staining is to be simulated. For the scanning of the slides, the scanner Pannoramic DESK from Sysmex© was used. Afterwards, a section of this image at a magnification of $100\times$ or $200\times$ was cut out for further evaluations. This image can then be opened with the program Fiji (Schindelin et al. 2012). The use of a sufficiently high magnification is necessary because otherwise the nuclei cannot be recognized unequivocally. Then, a binary mask was created with the "Trainable Weka Segmentation" plugin, generated by Arganda-Carreras et al. (2017), shown in Fig. 5b. The image is divided into nuclear areas and areas free of nuclei, e.g., cytoplasm or interstitium or void spaces with this mask. The binary mask can now be saved and loaded into MATLAB. With the MATLAB function isocontour (Kroon 2021), the isocontour lines were calculated and the information was saved in form of the isocontour lines and the points that make up those lines (Fig. 5c). For the creation of the actual FE mesh for the projection, the program Gmsh (Geuzaine and Remacle 2009) was used. Therefore, the information of the isocontour lines and points is exported from MATLAB in a syntax, which Gmsh can read. After that, the meshing algorithm from Gmsh can be used to create the FE mesh (Fig. 5d). The different colors are just a visual presentation of (i) the different nuclei (multiple colors) and (ii) the non-nuclear components of the tissue (green).

Summary of the assumptions

The assumptions made to create the model can be summarized as follows:

1. The concentration of the dye on the surface of the histological slide is constant for the time period under consideration.
2. The reaction of the dye with the tissue is a reversible equilibrium reaction.
3. The diffusion in the xy-plane can be neglected.
4. The number of binding sites for the dye in the tissue is fixed ergo there is a c_b^{max}.
5. The thickness of the section and the diameter of the cell nuclei inside the histological section are usually in the same range.
6. The cell nucleus has the shape of a sphere.
7. The temperature T inside the dye basin is constant.

8. The diffusion coefficient D inside the slide is isotropic and homogeneous.
9. There exists only one cell nucleus over the height of the slide.

These assumptions apply only to the staining of liver tissue with hemalum. However, for tissues with similar structure, the basic procedure could be the same. The staining of other tissues or the use of other dyes and staining techniques will be part of future work.

Results

For validation and comparison of the proposed model, frozen sections of the liver were stained with Mayer's hemalum for different periods of time. Sections from these slides are shown in Supplementary Fig. 4. Two individual slides were prepared for each of the six staining times, $t = 0s, 13s, 25s, 50s, 100s, 200s$. For simplicity, it was assumed that the staining within a nucleus is evenly distributed. Therefore, it is sufficient to determine the RGB intensities for the central point of the nuclei. For each time, one of the two prepared slides was selected. Three different sections were selected from this slide and investigated at a $200\times$ magnification to guarantee a sufficient pixel number per nucleus.

Preliminary studies have shown that the magnification should not be too small. With a lower magnification, more cell nuclei can be considered, but as a consequence the number of pixels per nucleus is too low. To counteract this, the $200\times$ magnification is chosen. Additionally, multiple sections were selected to determine the mean values of the RGB intensities for each staining time point. For these sections, the geometry was approximated and the central point of each cell nuclei be determined. With the coordinates of the central points, the RGB intensities from the sections could be computed.

The binary mask stores the information of whether pixels belong to a nucleus or not. However, this mask may also contain artifacts in which only part of a nucleus was detected or a certain part of the cytoplasm fell into the class of nuclei (Supplementary Fig. 3). In order to filter out incorrectly detected nuclei, the area of each nucleus was computed. Afterwards, only those "nuclei" whose area was larger than $A = 12(\mu m)^2$ were processed. The exact value used for the threshold of the area is chosen arbitrarily. The main reason for the need of a threshold value is to filter out small artifacts that are clearly not nuclei (Supplementary Fig. 3).

Figure 6 shows the result of combining the two procedures: (i) filtering out the artifacts and (ii) determining the RGB intensities via the central points of the nuclei. Three different sections were evaluated per staining time

Fig. 6 Determination of mean RGB values for different time points of tissue section staining. Three histological slides from which sections were taken are depicted in the image. For each staining time point, a different histological section was evaluated in the experiments
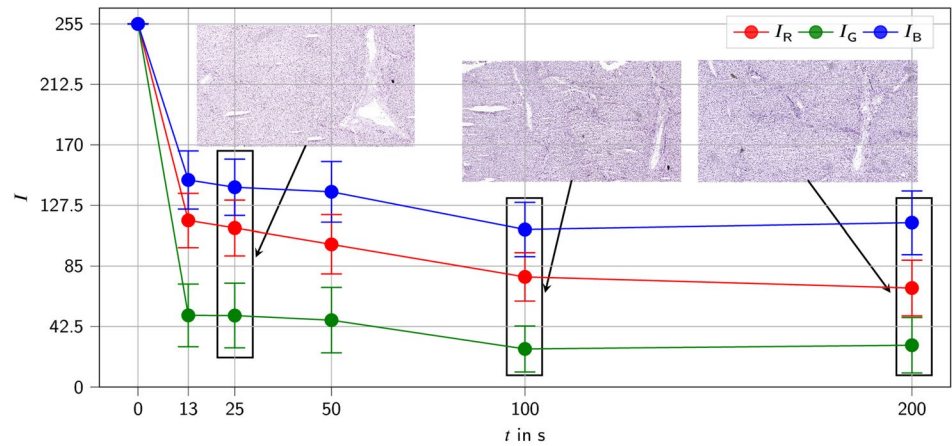
corresponding to total of approximately 2400 nuclei per staining time.

The error bars defined by the standard deviation are almost the same for each time and quite large. To give an overview of the variation in the RGB intensities that are computed by this method, box plots of the RGB intensities are shown in Supplementary Fig. 6. It can be seen that the number of outliers is quite different for the different sections. However, the number of outliers is small compared to the total values studied. Furthermore, the mean values were fairly similar for all three sections (of the respective histological slide) per staining time.

In Fig. 6 it can be seen that after a staining time of only 13 s , the nuclei already exhibited a very intense staining. For longer staining times, the intensities increased much

slower and after $t = 100$s they did not increase anymore (Fig. 6). Therefore, it can be concluded that between $t = 50$s and $t = 100$s an equilibrium has been reached. The mean RGB intensities determined depend on which of the two slides is used for each stain (Supplementary Fig. 2).

The quality of the data can be improved by choosing more similar tissue samples. Another possibility is to remove the same tissue sample at different staining times, which is experimentally demanding. However, it should be noted that this does not impact the simulation results, because only two individual slides are used for the fitting: the first one at $t = 25$s (in disequilibrium) is used for calculating the diffusion coefficient and the rate constants, the second one (in equilibrium) at $t = 100$s is used for the calculation of
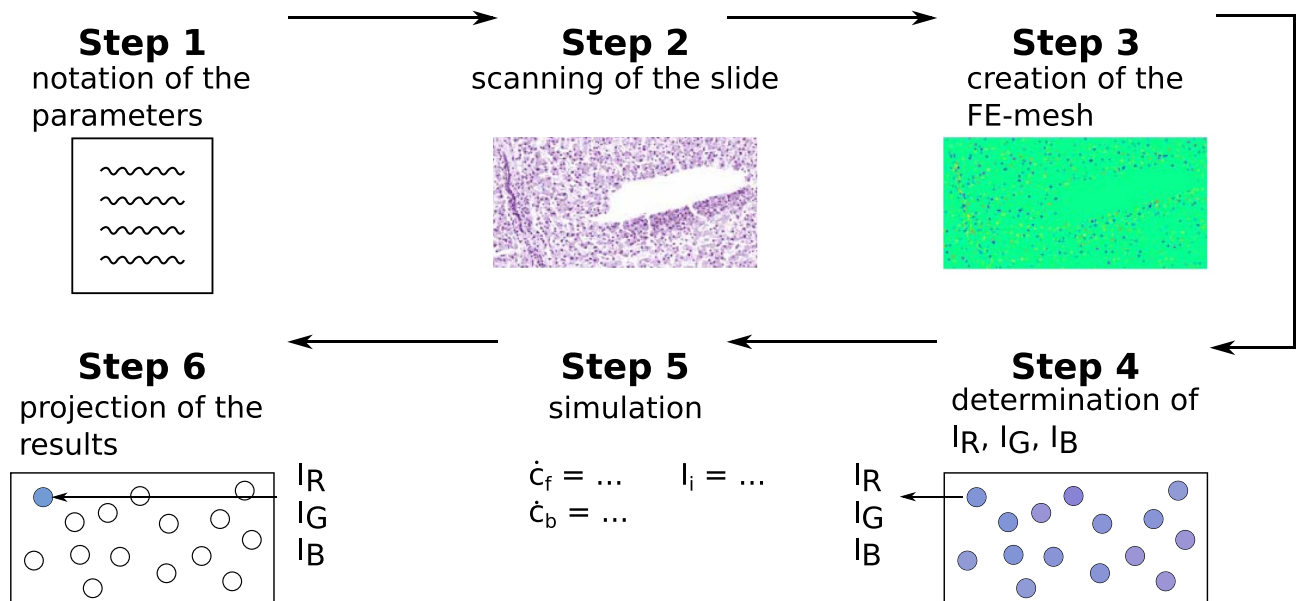

Fig. 7 Flowchart of the single steps needed to conduct the simulation for the virtual staining of a histological slide. According to the initial conditions for c_b and c_f, both concentrations are set to 0 mol m^{-3} at $t = 0$ s, cf. Section "Initial conditions and boundary conditions"

extinction coefficients as well as the maximal dye concentration, which the tissue can bind.

The typical workflow for performing a virtual staining is shown in Fig. 7. First, the parameters used for the creation of the histological slide, like the thickness of the tissue section and the dye used, should be written down. After that, the already-stained slide for which the staining with another dye or for another time span should be simulated is scanned. Based on this scanned slide, a FE mesh is created (see "Finite element mesh from a scanned slide") with which the geometry is determined.

For each nuclei, an individual c_b^{max} is calculated. For that purpose, first the RGB intensities of each nucleus were determined. Subsequently, a series of test calculations were performed with the presented model, in which different values for c_b^{max} were given and the RGB intensities were calculated. This served as a calibration series for the relationship between the local c_b^{max} and the color intensity.

For the determination of the maximal number of binding sites, it is important to use a stained slide. This is also the reason why for this model the starting point is a stained slide and not an unstained one. Another requirement is that the dye with which the physical slide was stained and the dye with which the virtual staining has to be done, have the same binding sites. If a method is found to determine c_b^{max} purely by geometrical features, a virtual staining could also be done on the basis of an unstained slide.

The RGB intensities obtained by the test calculations are then combined with the respective concentration c_b^{max} used as the material parameter to form an interpolation table. This allows the definition of c_b^{max} based on a RGB intensity of a nucleus. Afterwards, the dye concentration can be computed with the Eqs. (2) and (3) and the resulting staining with the Eqs. (6), (7), and (8) for each cell nuclei individually. Finally, the calculated RGB intensities can be projected onto the FE mesh generated in step 3 (Fig. 5). Now, the three intensities are combined to form the respective color. The comparison between numerically simulated and scanned/measured color is shown in Supplementary Fig. 5.

For the validation of the proposed model with the experimental data, the steps described above in the workflow are performed for a tissue section which was stained for $t = 100$ s. With that, the geometry and c_b^{max} are fitted. The distribution of the free and bound dye concentration c_f and c_b is shown and described in Supplementary Information.

The RGB intensities were calculated for a staining time of 200 s. For the real, physical slides, a different slide was used for each staining time (Supplementary Fig. 4). Therefore, it is not possible to directly compare point-wise (or nucleus-wise) the results of the projected RGB intensities with the real, physical sections. A visual comparison can only be done qualitatively as will be reported below. For a quantitative comparison of the simulation with the experimental

results, only a global comparison is possible. The assumption was made that if the global comparison, that is over all the nuclei, is good, the local comparison for the single nuclei would also be good. For a global comparison, the average of the RGB intensities of all nuclei was taken for each simulated second. The simulation results are plotted together with the mean values and error bars of the experimental data versus time in Fig. 8. The simulation results show a fast decrease of the RGB intensities in the first seconds followed by a plateau leading to the final steady-state value (equilibrium value) which is unique for each color. After a staining time of $t = 50$ s the equilibrium seems to be reached. After a staining time of $t = 42$s, 95% of the maximal staining is achieved for the simulated data. For $t = 25$s and $t = 100$s the simulation and the experimental data are in excellent agreement. This was to be expected, as the data of the experiments for this time point was used to fit the material parameters used in the equations.

For the other time points investigated, the simulation results were always within the range spanned by the error bars with the exception of I_G for $t = 13$s and I_B for $t = 50$s. A possible factor for the deviation of the simulation from the experimental results could be the fact that the thickness of the experimental histological sections could be different from the assumed thickness $h = 6\mu m$. Please note that $h = 6\mu m$ is the height set on the microtome. Since this is a highly precise work step on inhomogeneous (biological) material, variations in the cutting result are quite typical. Estimating the actual section thickness would be one way of adding a correction factor. So far, this has not been possible in standardized applications.

Collectively, our results suggest that the proposed model is well suited to simulate the staining of histological sections. In Supplementary Fig. 6, the box plots of the
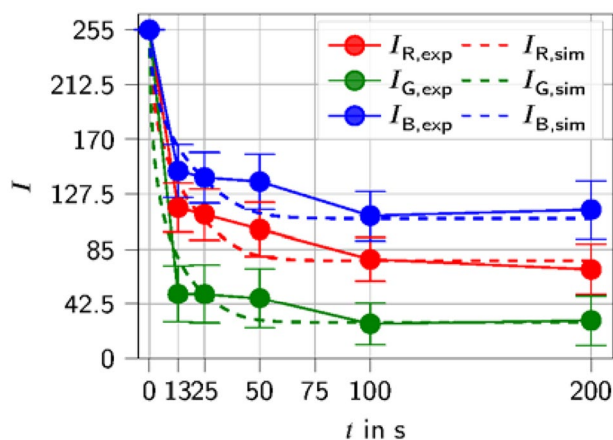

Fig. 8 Comparison of the mean values of the RGB intensities of the real, experimental investigated stained slides (exp) with the numerically simulated staining (sim). For the calibration of the geometry and c_b^{max} the used section is depicted in Supplementary Fig. 5

simulated RGB intensities of the nuclei are depicted. The number of outliers and their value is much lower than for the experimental data. It can be observed that for longer staining times the distribution of the RGB intensities increases.

However, the goal of the proposed model is not to create a diagram as seen in Fig. 8, but to create 2D images of virtually stained histological sections that can be evaluated by an investigator or a machine learning algorithm for cancer detection. In Supplementary Fig. 5, a comparison of simulated staining with the experimental results for the different staining times is shown. It should be noted that only the staining of the nuclei was simulated in the present work. In addition to the stained nuclei (darker purple), the experimental data also include sections of stained cytoplasm (lighter purple), see Fig. 1.

For the staining time of $t = 13s$, many of the virtually stained nuclei were not as intensely stained as the experimental one (see also Fig. 8). For the later time points, a better agreement between the results of the virtual and the actual staining was found. The variation in intensities of the virtually stained nuclei is larger than in the experimental data (Supplementary Fig. 5). This could be a sign that the process of choosing a c_b^{max} for each nucleus should be improved. Overall, the difference between the simulated staining of the nuclei and the experimental data is within a tolerable range.

Discussion and outlook

In the present work, a model capable of performing virtual staining of cell nuclei in tissue sections is presented. With the proposed model, we lay the foundation for a histological analysis that can be adapted to the personal habits/experiences of the histologists. It also allows the generation of a much wider variety of training data sets for machine learning methods due to the possibility to standardize the images. Thus, our method allows the application of machine learning-based cancer detection algorithms to slides that are stained with another procedure.

The transport and binding of the dye was computed using a one-dimensional FE simulation. Afterwards, the resulting RGB intensities were determined and the intensities were projected onto a two-dimensional mesh based on a scanned histological slide. When the used material parameters were fitted with the data from two different time points, the obtained curves were generally within the range spanned by the error bars. This means that the model is able to accurately simulate the staining of the nuclei with hemalum over time. In a purely visual comparison, the two-dimensional simulation results are similar to the experimental data. However, there are still small differences, especially for shorter staining times.

With our model, a user can virtually (re-)stain cell nuclei of a histological section from liver tissue, or tissues with similar properties using the same or a different dye. A prerequisite for the applicability of the model is that the cell nuclei of the tissue under investigation form the same bonds with hemalum, which can be assumed at least for most human tissues. For tissue sections of interest, the progress of the RGB intensities at different time points can be visualized. For that, the geometry of the nuclei is approximated. Often a tumor cell is not or not only characterized by a different staining intensity compared to a normal parental cell but also by a larger diameter, which can be detected by the described model. Thus, the investigator or the artificial intelligence model analyzing the virtual staining, perceives this and can make a decision.

An important aspect for further research concerning this topic is the experimental determination of the material parameters like the diffusion coefficient D or the rate constants k_{on} and k_{off}. An expansion of the current model, which simulates the staining of nuclei, to include the staining of the cytoplasm is another important aspect. Only the combination of nuclear and cytoplasmic staining will potentially permit a truthful diagnosis.

Declarations

References

Alnæs MS, Blechta J, Hake J, Johansson A et al (2015) The FENICS project version 15. Arc Numerical Software 3(100):9–23

Arganda-Carreras I, Kaynig V, Rueden C, Eliceiri KW et al (2017) Trainable WEKA segmentation: a machine learning tool for microscopy pixel classification. Bioinformatics 33(15):2424–2426

Ari A, Hanbay D (2018) Deep learning-based brain tumor classification and detection system. Turk J Electr Eng Comput Sci 26(5):2275–2286

Bejnordi BE, Litjens G, Timofeeva N, Otte-Höller I et al (2015) Stain specific standardization of whole-slide histopathological images. IEEE Trans Med Imaging 35(2):404–415

Bettinger C, Zimmermann H (1991) New investigations on hematoxylin, hematein, and hematein-aluminium complexes. II. Hematein-aluminium complexes and hemalum staining. Histochemistry 96(3):215–228

de Cea MVS, Nie Y (2018) Single stain normalization for IHC whole slide images. Med Imaging 2018 Digit Pathol 10581:1058104

de Souza SMAGU, de Souza DP, da Silva EAB, de Souza AAU (2007) Modelling of the dyeing process of packed cotton threads using reactive dyes. Transp Porous Media 68(3):341–363

Dong H, Yang G, Liu F, Mo Y, Guo Y (2017) Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: Annual conference on medical image understanding and analysis. Springer, Cam, pp 506–517

Falk M, Ynnerman A, Treanor D, Lundström C (2018) Interactive visualization of 3D histopathology in native resolution. IEEE Trans vis Comput Graph 25(1):1008–1017

Ferlay J, Ervik M, Lam F, Colombet M et al. (2020). Global cancer observatory: Cancer today. Lyon: Int. agency for research on cancer. https://www.who.int/news-room/fact-sheets/detail/cancer. Accessed 3 Aug 2021

Ferus-Comelo, M. (2002). Control of the adsorption of dyes on cotton. Doctoral dissertation, University of Leeds.

Fibich G, Hammer A, Gannot G, Gandjbakhche A, Gannot I (2005) Modeling and simulations of the pharmacokinetics of fluorophore conjugated antibodies in tumor vicinity for the optimization of fluorescence-based optical imaging. Lasers Surg Med 37(2):155–160

Gavrilovic M, Azar JC, Lindblad J, Wählby C et al (2013) Blind color decomposition of histological images. IEEE Trans Vis Comput Graph 32(6):983–994

Geuzaine C, Remacle J-F (2009) Gmsh: a 3D finite element mesh generator with built-in pre-and post-processing facilities. Int J Numer Methods Eng 79(11):1309–1331

Giuliano AE, Hunt KK, Ballman KV, Beitsch PD et al (2011) Axillary dissection vs no axillary dissection in women with invasive breast cancer and sentinel node metastasis: a randomized clinical trial. JAMA 305(6):569–575

Goldstein D (1980) A microdensitometric method for the analysis of staining kinetics. J Microsc 119(3):331–343

Haroske G, Baak J, Danielsen H, Giroud F et al (2001) Fourth updated ESACP consensus report on diagnostic DNA image cytometry. Anal Cell Pathol 23(2):89–95

Heid H, Reith J (2010) Malerfachkunde, 5th edn. Springer Fachmedien Wiesbaden

Hiddemann W, Bartram CR (2010) Die Onkologie, 2nd edn. Springer

Hollon TC, Pandian B, Adapa AR, Urias E et al (2020) Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nat Med 26(1):52–58

Irion G, Ochsenfeld L, Naujok A, Zimmermann H (1993) The concentration jump method. Histochemistry 99(1):75–83

Jones TL, Byrnes TJ, Yang G, Howe FA, Bell BA, Barrick TR (2015) Brain tumor classification using the diffusion tensor image segmentation (d-seg) technique. Neuro-oncology 17(3):466–476

Kather JN, Pearson AT, Halama N, Jäger D et al (2019) Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 25(7):1054–1056

Kroon D-J (2021) Isocontour. MATLAB Central File Exchange: https://www.mathworks.com/matlabcentral/fileexchange/30525-isocontour. Retreived 2 Feb 2021

Lang G (2006) Histotechnik: Praxislehrbuch für die biomedizinische Analytik. Springer

Lin S (1992) Concentration-dependent diffusion of dye in reactive dyeing systems. J Appl Polym Sci 44(10):1743–1749

Menning JDM (2021) Kontinuumsbasierte Modellierung und Simulation von Diffusions- und Reaktionsprozessen bei der Färbung histologischer Schnitte. Diplomarbeit, Institut für Festkörpermechanik, Technische Universität Dresden

Mulisch M, Welsch U (2010) Romeis - Mikroskopische Technik, 18th edn. Spektrum Akademischer Verlag Heidelberg

Prentø P (2001) A contribution to the theory of biological staining based on the principles for structural organization of biological macromolecules. Biotech Histochem 76(3):137–161

Prentø P (2009) Staining of macromolecules: possible mechanisms and examples. Biotech Histochem 84(4):139–158

Rassow J, Hauser K, Netzker R, Deutzmann R (2012) Duale reihe biochemie, vol 3. Thieme, Stuttgart

Reddy M, Lee G, McGregor R, Jasper W (1995) Modeling of the batch dyeing process. Proc of 1995 American control conference-acc'95 3, 2180–2184.

Ruifrok AC, Johnston DA et al (2001) Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 23(4):291–299

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V et al (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9(7):676–682

Sharma K, Kaur A, Gujral S (2014) Brain tumor detection based on machine learning algorithms. IJCA 103(1):7–11

Siedel I, Zimmermann H (1995) Relaxation technique for kinetic measurements of dye accumulation at the mitochondria of hela cells in situ. J Photochem Photobiol B Biol 27(3):233–242

Suvarna SK, Layton C, Bancroft JD (2019) Bancroft's theory and practice of histological techniques, 8th edn. Elsevier Health Sciences

Van der Laak JA, Pahlplatz MM, Hanselaar AG, de Wilde PC (2000) Hue-saturation-density (HSD) model for stain recognition in digital images from transmitted light microscopy. Cytometry A 39(4):275–284

Veuthey T, Herrera G, Dodero VI (2014) Dyes and stains: from molecular structure to histological application. Front Biosci 19(1):91–112

Winzek C, Baumgärtel H (1988) Staining kinetics in single cells. Part II. diffusion processes inside the cell. Histochemistry 90(1):79–83

Winzek C, Plieninger P, Baumgärtel H (1987) An improved method to investigate staining kinetics in single cells. Histochemistry 86(4):421–426