Cell and Tissue Research (2019) 375:371–381
https://doi.org/10.1007/s00441-018-2911-1

REGULAR ARTICLE

Digital pathology imaging and computer-aided diagnostics as a novel tool for standardization of evaluation of aganglionic megacolon (Hirschsprung disease) histopathology

Florian Schilling[1,2] · Carol E. Geppert[2] · Johanna Strehl[2] · Arndt Hartmann[2] · Stefanie Kuerten[1] · Axel Brehmer[1] · Samir Jabari[1,2]

Received: 23 March 2018 / Accepted: 15 August 2018 / Published online: 3 September 2018

Abstract

Based on a recently introduced immunohistochemical panel (Bachmann et al. 2015) for aganglionic megacolon (AM), also known as Hirschsprung disease, histopathological diagnosis, we evaluated whether the use of digital pathology and 'machine learning' could help to obtain a reliable diagnosis. Slides were obtained from 31 specimens of 27 patients immunohistochemically stained for MAP2, calretinin, S100β and GLUT1. Slides were digitized by whole slide scanning. We used a Definiens Developer Tissue Studios as software for analysis. We configured necessary parameters in combination with 'machine learning' to identify pathological aberrations. A significant difference between AM- and non-AM-affected tissues was found for calretinin (AM 0.55% vs. non-AM 1.44%) and MAP2 (AM 0.004% vs. non-AM 0.07%) staining measurements and software-based evaluations. In contrast, S100β and GLUT1 staining measurements and software-based evaluations showed no significant differences between AM- and non-AM-affected tissues. However, no difference was found in comparison of suction biopsies with resections. Applying machine learning via an ensemble voting classifier, we achieved an accuracy of 87.5% on the test set. Automated diagnosis of AM by applying digital pathology on immunohistochemical panels was successful for calretinin and MAP2, whereas S100β and GLUT1 were not effective in diagnosis. Our method suggests that software-based approaches are capable of diagnosing AM. Our future challenge will be the improvement of efficiency by reduction of the time-consuming need for large pre-labelled training data. With increasing technical improvement, especially in unsupervised training procedures, this method could be helpful in the future.

Keywords Calretinin · Digital pathology · Hirschsprung disease · Immunohistochemistry · Machine learning

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00441-018-2911-1) contains supplementary material, which is available to authorized users.

✉ Samir Jabari
samir.jabari@fau.de

Florian Schilling
florian.schilling@studium.fau.de

Carol E. Geppert
carol.geppert@uk-erlangen.de

Johanna Strehl
johanna.strehl@uk-erlangen.de

Arndt Hartmann
arndt.hartmann@uk-erlangen.de

Stefanie Kuerten
stefanie.kuerten@fau.de

Axel Brehmer
axel.brehmer@fau.de

[1] Institute of Anatomy and Cell Biology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Krankenhausstraße 9, 91054 Erlangen, Germany

[2] Institute of Pathology, Friedrich-Alexander-Universität Erlangen-Nürnberg, Krankenhausstraße 9, 91054 Erlangen, Germany

Introduction

Aganglionic megacolon (AM) known as Hirschsprung disease is a congenital malformation displaying aganglionosis in the human colonic myenteric and submucosal plexus (Butler Tjaden and Trainor 2013). Approximately 1 of 5000 live-born infants suffers from AM with a predominance in males (sex ratio 4:1) (Badner et al. 1990; Bodian and Carter 1963). Short-segment AM is the most common version and affects the intestinal segments between the rectum and the splenic flexure, while long-segment AM continues orally to this flexure (Burkardt et al. 2014). Straight after birth, AM generally presents with deferred meconium as clinical symptom. Approximately less than 10% of affected neonates pass meconium in first 24 h of life. According to this, any fully developed infant with more than 24 h and any premature infant with about 7–8 days delayed meconium is suspected of being affected with AM (Bekkali et al. 2008; Clark 1977). Most Hirschsprung disease patients develop dysmotility and constipation symptoms, which begin postnatally and are usually detected within the first three months. A minority of 1% is first diagnosed as adults (Barnes et al. 1986) and can lead to possible complications (Murphy and Puri 2005). Therapy options implicate intestinal surgery with removal of the aganglionic segment and renewal of the intestinal permeability one year after birth followed by intermediate therapy with colostomy (Duhamel 1960; Martin and Altemeier 1962).

Several methods exist to confirm the diagnosis of Hirschsprung disease. However, neither radiological imaging of the transition zone nor anorectal manometry with 90% concordance is able to confirm AM in neonates (Lopez-Alonso et al. 1995; Pratap et al. 2007). Suction biopsy is the gold standard in AM diagnosis with about 99% sensitivity, being safe, cheap and easy to apply (Andrassy et al. 1981; Dobbins and Bill 1965; Harjai 2000). In cases of inconclusive histological diagnosis, a standard transmucosal biopsy of the rectum is additionally needed (Holland et al. 2010). It turns out that the location of biopsy is decisive. Due to a hypoganglionic zone, only biopsies 2–3 cm above the pectinate line provide diagnostically conclusive results. The zone below is characterized by physiological absence of neurons so that false positive findings may occur (Aldridge and Campbell 1968). The histological aspect of AM-induced aganglionosis includes hypertrophic preganglionic nerve fibres with an excess of acetylcholine. The common diagnostic method is a combination of acetylcholinesterase staining and haematoxylin eosin (H&E) staining of frozen sections obtained by suction biopsy of rectal tissue (Meier-Ruge et al. 1972). Additionally, immunohistochemical staining can aid histopathological diagnostics. There are exhaustive lists of nerve fibre markers such as S100, beta-tubulin (S100β), glial fibrillary acid protein (GFAP), nerve growth factor receptor (NGFR) and cell body markers like calretinin, MAP2, peripherin, NeuN, neuron-specific enolase, nNOS and RET oncoprotein which were used for labelling acridine orange ribosomal fluorochrome, as an older technique which were used for labelling (Barshack et al. 2004; Guinard-Samuel et al. 2009; Holland et al. 2010; Joosten et al. 1989; Kakita et al. 2000; Kapur 2014; Kapur et al. 2009; Sarnat et al. 1985; Taguchi et al. 1985; Volpe et al. 2013; Yang and Oh 2013; Yanlei et al. 2011).

Nevertheless, a reliable statement can only be made by time-consuming examination of a multitude of sections by an experienced pathologist (Kapur and Kennedy 2012; Martucciello 2008; Qualman et al. 1999; Szylberg and Marszalek 2014). More innovative approaches have not yet been included in the perioperative or intraoperative diagnosis. Over the last 10 years, the technological progress and advance of whole slide digital scanners has led to an increasing digitalization of histopathological slides and conclusively software-based pathologist assisting systems. This progress is based on ‘machine and deep learning’, a technique that is able to process big amounts of data defined as training sets using a neural network algorithm (Madabhushi and Lee 2016). Here we aimed to evaluate a three-step approach. First, we digitized immunohistochemically stained slides. Second, we performed morphometric analysis. Third, we evaluated the morphometric data via machine learning models. Although histopathological analysis also allows for differential diagnosis of other diseases, our focus was exclusively kept on discriminating between Hirschsprung and non-Hirschsprung cases. To our knowledge, this is the first attempt to train a neuronal classifier to discriminate Hirschsprung from non-Hirschsprung cases.

Materials and methods

Tissue processing, immunohistochemistry and digitalization

We used specimens of 27 patients who were suspected to be AM positive between 2013 and 2016. The patients’ ages ranged from 31 days to 382 months (31.83 years) (median age 61.98 months (5.17 years)). We used paraffin-embedded and 4% formalin-fixed (FFPE) material stored in the archive of the Institute of Pathology, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany. The use of the FFPE materials was approved by the local ethics committee (approval number: 85_12b, date 04/19/2012). The Declaration of Helsinki was followed strictly.

In total, we collected 93 tissue blocks from 14 female and 13 male patients originating from nine surgical specimens (colon/rectum resections) and 84 suction biopsies. The biopsies were obtained from 22 of the 27 patients resulting in a total number of 31 specimens. Based on these tissue samples, 75 slides were stained for calretinin, 86 slides for MAP2, 58 slides for S100β and 88 slides for GLUT1. By using the

immunohistochemical panels described by Bachmann et al. (2015) for diagnosing AM, 12 of 31 specimens fulfilled the criteria of AM (median age of the patients 38.72 months; gender: five females and seven males). One of these 12 specimens was described as short-segment AM (median patient age 81.21 months; gender: male), while two specimens were first not clearly diagnosed (median age of the patients 48.57 months; gender: two males) and needed closer consideration. Negative diagnosis was stated for 19 specimens (median age of the patients 71.68 months; gender: 11 females and eight males). With regard to the count of 27 patients, 19 were not affected (11 females and eight males/16 biopsies, three resections with one re-surgery and eight were diagnosed as AM positive (three females and five males/six biopsies and two resections)). After immunohistochemistry the above-mentioned slides were digitalized using a digital slide scanner in bright-field mode (Pannoramic Flash 250 for bright-field, 3D-Histech AG, Budapest, Hungary).

Quantitative analysis using digital pathology programmes

We used Definiens Developer (Tissue Studios) to analyse each slide. Tissue Studios allows the configuration of own solutions aligned to parameters necessary to define specific cellular structures. The software enables the analysis of whole areas of interest in various slides. The so-called Analysis Builder allows to adjust settings step by step to refine the results and can be adopted. For further in-depth specifications, we refer to Electronic Supplementary Material, Table S1. Briefly, we defined a Training Data Set for each immunohistochemical staining by importing 12 slides of each staining to work with simultaneously. In this way, a better adaptation and generalization could be achieved in terms of variations in staining, scanning, biological variance and manifestation of disease.

First of all we adjusted the 'general settings' containing image information (pixel/resolution) using the metadata from the image file and staining information (colouration and stained structure). Since we intended to analyse the submucosa exclusively, a region of interest (ROI) had to be defined by manually setting marks. Unlike random selection of high-power fields, this enabled submucosal tissue evaluation in total. This 'ROI detection' feature does not refer to spectral information and therefore allows defining areas in any image. For analysis of the selected ROI, subsets and their magnification had to be defined. Subsequently, to detect and classify different stains, we used 'marker area detection'. In the cases of calretinin and MAP2, we defined 'IHC marker stain (brown chromogen)' thresholds to identify neurons within ganglia and nerve fibres in the case of S100β staining. In principle, the software applies a 'stain isolation' algorithm based on chromogen or haematoxylin-eosin to discriminate tissues. Based on, e.g. the area and the length-to-width ratio, the 'marker area classification exclusion' action enables the software to eliminate false positive artefacts appearing as 'IHC marker stain (brown chromogen)' areas by adjusting thresholds of the above-mentioned features (Fig. 1). For nerve fibre identification, we used the 'vessel detection' feature analysing GLUT1-stained slides. To rule out deviation through mistakenly classified vessels, adapted ROI selection had to be specified to nerve fibres exclusively by manual selection and not to submucosa in total. As Hirschsprung disease shows broadened perineuria of nerve fibres, the 'vessel classification' action enables the measurement of the vessel wall thickness, which proved to be equivalent to the thickness of perineuria and determining their quantity. We describe this as 'nerve recognition and measuring function' in the following paper. Equally to 'marker area classification', we had to eliminate false positive artefacts based on thickness and area by adjusting thresholds (Fig. 1). By using the 'stain isolation' algorithm, the software was able to separate nerve fibre-like structures and separate them from not relevant tissue.

By configuring a standard 'default export' and an additional 'custom export' for results (see Electronic Supplementary Material, Table S2), we received statistics for each slide or on purpose for each ROI (Definiens 2015).

Statistical analysis

Statistics were performed using Python 3.5 and the Scipy.stats package as well as the statsmodels Python module. p values < 0.05 were considered significant. The exported parameters mentioned above were tested using independent Mann–Whitney U test comparing non-AM controls with AM data. We additionally tested whether there was a difference for suction biopsies and resections within the groups by again using the independent Mann–Whitney U test.

Machine learning

Using the scikit-learn (0.19.1) framework, we evaluated different machine learning models and algorithms via grid search. We finally built a voting ensemble classifier containing a logistic regression classifier, a support vector machine, a decision tree classifier and a random forest classifier. The applied parameters and hyperparameters can be seen in Table S4 of the Electronic Supplementary Material. For data pre-processing, we normalized the parameter values and rescaled them to the range from 0 to 1. We split our data into one training, development and test set. Care was taken that all sets came from the same distribution by preserving the percentage of samples for each class. This was performed by applying the scikit-learn StratifiedShuffleSplit method. The training set contained 56% of the cases, the development set 19% and 25% were left for the test set. Fitting the classifier was

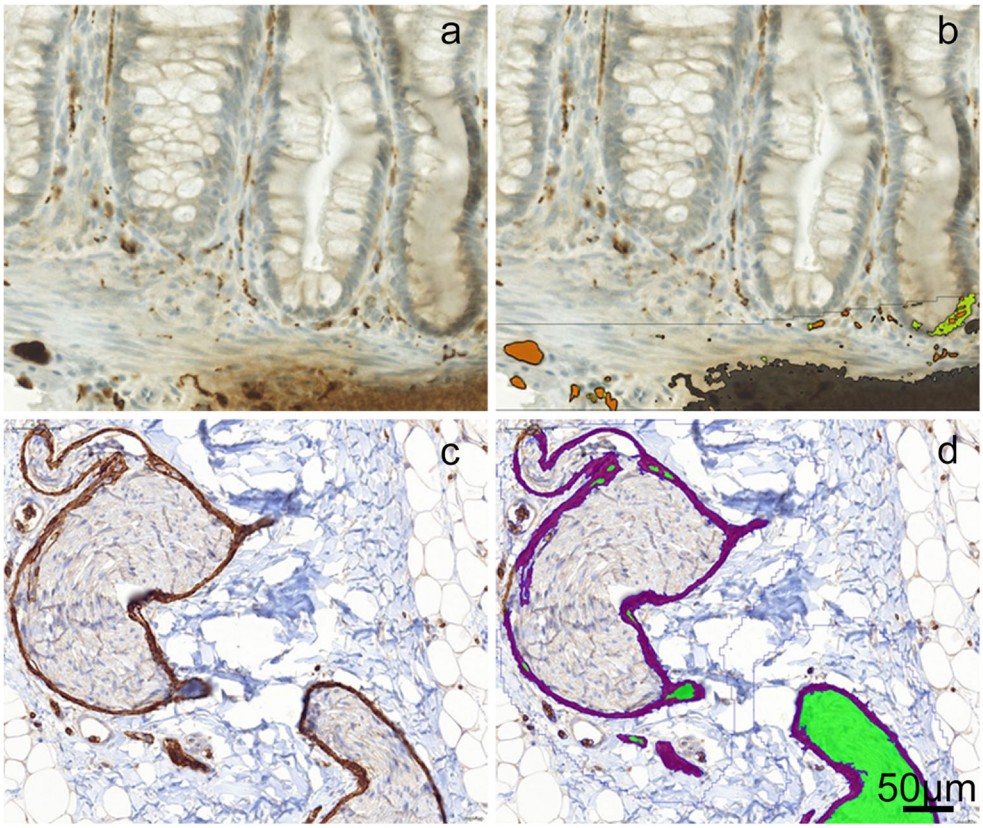


Fig. 1 Automated detection and classification of trained features by Definiens pathology software. **a–b** Adjusted and algorithm-based detection and classification of 'IHC ,marker stain (brown chromogen)' areas in calretinin-stained submucosal tissue of the human large intestine. False positive artefact areas were excluded and covered in grey, yellow and orange resemble detected ganglion cells (**a**: raw input, **b**: adjusted output). **c–d** Algorithm-based, automated detection, classification and measurement of nerve fibre perineuria applying the 'vessel detection and classification action' in GLUT1-stained submucosal tissue of the human colon. Violet resembles detected nerve perineurium and green lumen (**c**: raw input, **d**: detected output). Digitalized using a digital slide scanner in bright-field mode (Pannoramic Flash 250 for bright-field, 3D-Histech AG, Budapest, Hungary) and processed by applying Definiens Developer Tissue Studios 4.3

performed on the training set. Then, all parameter tuning was performed on the development set. The test set was evaluated only at the end of parameter tuning. We resampled the minority class to match the majority class. Finally, we constructed our model containing the above-mentioned classifiers and evaluated different performance metrics such as accuracy, recall and area under the curve. We thereby applied 10-fold cross-validation. In a second approach, we additionally split up the AM and non-AM groups into two subgroups consisting of suctions biopsies and resections.

Results

The Definiens software classifies the data by different parameters. After a preselection of all features, which are shown in Electronic Supplementary Material, Table S3, we received a number of potential parameters with relation to significance in diagnosing AM. The supposed features have been selected and computerized by an algorithm (Table 1).

The automatic software-based process reveals significance for parameters applied in calretinin and MAP2 stainings but not in S100β and GLUT1 stainings.

Calretinin- and MAP2-stained neurons and ganglia

Calretinin- and MAP2-stained submucosal neurons were identified by applying the software as described above. Figure 2a–d shows digitized immunohistochemically stained slides of calretinin and MAP2 in non- and AM-affected tissues. The difference between normal tissue and aganglionosis is shown in Fig. 2a, b for calretinin and in Fig. 2c, d for MAP2. The configured algorithm detected those stained areas by using an 'IHC marker stain (brown chromogen)' filter, adjusted to criteria we have set in advance. Figure 2a, b shows the automated detection of neurons and ganglia for calretinin and Fig. 2c, d for MAP2. Areas, which were identified as neuronal structures, were stained in orange and red. The threshold of 'IHC marker stain (brown chromogen)' was 0.57 for calretinin and 1.11 for MAP2. The configured solution for calretinin and MAP2 shows a high predictive accuracy with regard to actual

Table 1 Preselection of parameters with supposed correlation to diagnosis. Our software approach classifies the results of data by various parameters. The ones with a supposed correlation to diagnosis have been preselected and computerized. The algorithm-based selection revealed '% area IHC marker stain (brown chromogen)' for calretinin and MAP2 to be significant referring to the diagnosis of AM (red). The assumed results of S100β ('% area IHC marker stain (brown chromogen)') and also GLUT1 ('median vessel wall thickness') showed no significant parameter related to AM-affected or healthy submucosal tissues of the human colon. A complete overview of all features computed by the software is shown in Electronic Supplementary Material, Table S3

Calretinin	% area IHC marker stain (brown chromogen)	Average brown chromogen intensity (IHC marker stain)	Area IHC marker stain (μm^2) (brown chromogen)	No. marker areas
MAP2	% area IHC marker stain (brown chromogen)	Average brown chromogen intensity (IHC marker stain)	Area IHC marker stain (μm^2) (brown chromogen)	No. marker areas
S100β	% area IHC marker stain (brown chromogen)	Average brown chromogen intensity (IHC marker stain)	Area IHC marker stain (μm^2) (brown chromogen)	No. marker areas
GLUT1	No. vessel	Average vessel wall thickness (μm)	Median vessel wall thickness (μm)	% vessel large

ganglia and false positive artefacts (Fig. 2a–d). The data for MAP2 and calretinin, which we acquired by the automated analysis of the software, contain many features (an almost exhaustive list can be seen in Electronic Supplementary Material, Table S3). Statistical analysis of the features showed a significant difference in the percentage of stained areas for calretinin, which accounted for 0.55% in AM cases and 1.44% in non-AM cases. The same was true for the MAP2 positive area percentage staining where the mean was 0.004% in AM and 0.07% in non-AM cases ($p = 0.0002$) (Table 2 and Fig. 3).

However, no difference could be found between resection and suction biopsy material (Table 3).

In addition, in MAP2-stained tissue, the percentage of the 'IHC brown area' in relation to the total submucosal tissue was lower than that in calretinin-stained samples (Table 2), corresponding to the results obtained by Bachmann et al. (2015).

S100β-stained nerve fibres

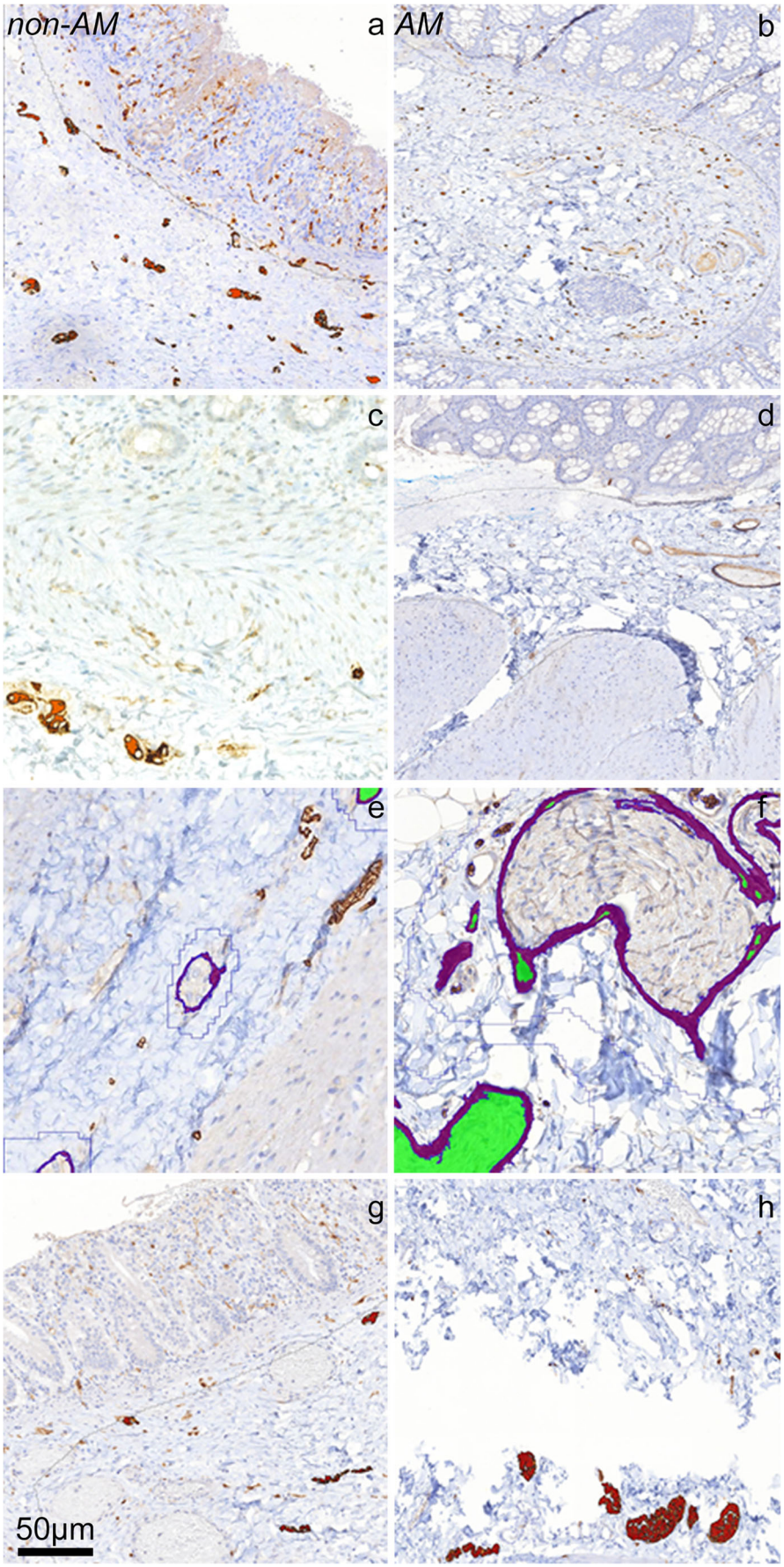
S100β marks submucosal nerve fibres. Figure 2g, h shows digitized images of S100β stainings in both non-AM- and AM-affected tissues. Nerve fibres in AM-positive tissue appeared thicker and denser compared to non-AM tissue of the human large intestine (Bachmann et al. 2015). As S100β marks nerve fibres in brown colour, the algorithm detects this colour by applying a brown filter similar to calretinin and MAP2. The automatically identified nerve fibre areas were stained in orange/red (Fig. 2g, h). The threshold of 'IHC marker stain (brown chromogen)' was 0.6. Actual nerve fibres were identified with a high predictive accuracy (Fig. 2g, h). In total, the thickness of nerve fibres in AM-positive tissue was larger than in non-affected tissue as shown in Table 2 and Fig. 3. The percentage of the stained area (resembling the nerve fibre thickness) for S100β accounted for 1.2% in AM cases and 0.9% in non-AM cases; however, this difference was not statistically significant. None of the parameters specified for S100β seen in Table 1 show significance. There was no concordance between digitally analysed nerve fibre thickness and AM-negative or AM-positive samples. We therefore conclude that nerve fibre thickness could be of less importance for the diagnosis of Hirschsprung disease which is in line with Bachmann et al. who also concludes that when in doubt neuronal presence or absence is of more importance.

GLUT1-stained nerve fibre perineuria

The immunohistochemical marker GLUT1 stains the perineuria of nerve fibres in submucosal tissue of the human large intestine and allows automated detection, analysis and measuring by our software-based approach. Figure 2e, f depicts the different shapes of nerve fibre perineuria in GLUT1-stained digitized slides of both non- and AM-affected tissues. The algorithm we adjusted detected perineuria using an 'IHC marker stain (brown chromogen)' filter and measured their thickness. Figure 2e, f shows the automatically detected part of the nerve fibre, which was stained in violet in non- and AM-affected submucosal tissues of the human colon. The green area was equal to the vessel lumen and was of no interest. The perineuria of nerve fibres were detected with a highly predictive accuracy by applying a threshold listed in Electronic Supplementary Material, Table S1. There is no GLUT1 parameter of the supposed selection in Table 1, which shows significance. The difference between thickness of perineuria in non-AM and AM-positive tissues was non-significant (Table 2 and Fig. 3). Overall, our results showed no concordance between analysed thicknesses of nerve fibre perineuria in AM-negative vs. AM-positive samples.

Machine learning

For machine learning purposes, we additionally labelled the features diseased or not diseased later on. All data acquired via software measurements were exported as .csv files and imported via pandas Python Data Analysis Library (v0.22.0) for further analysis. The features extracted were first preprocessed for all markers. We normalized the data to a unit norm and scaled the data to unit variance. We then resampled the minority class to match the majority class. Afterwards, we used scikit-learn framework to build our model. The performance of the classifier was measured via sensitivity,


non-AM
a
AM
b
c
d
e
f
g
h
50μm

◀ **Fig. 2** Software-based automated recognition and analysis of ganglion cells and nerve fibres in human submucosa of the colon. **a** Calretinin. Software-based detection of 'IHC marker stain (brown chromogen)'-stained neurons in normally innervated tissue. Orange areas resemble detected ganglion cells. **b** Calretinin. Characteristic aganglionosis detected by the software, signifying AM. Orange areas resemble detected ganglion cells. **c** MAP2. Software-based detection of 'IHC marker stain (brown chromogen)'-stained ganglia in normally innervated tissue. Orange resembles positive ganglion cells. **d** MAP2. Characteristic software detected aganglionosis, implying AM. Orange areas resemble detected ganglion cells. **e** GLUT1. Automated nerve fibre analysis in normally innervated tissue. Violet areas resemble nerve perineuria and green area resembles lumen. **f** GLUT1. Software-based detection of nerve fibres perineuria in AM-affected tissue. Thickened perineuria of nerve fibres compared to AM-negative tissue. Violet areas resemble nerve perineuria and green areas resemble lumen. **g** S100β. Automated 'IHC marker stain (brown chromogen)' nerve fibre detection in normally innervated tissue. Red areas resemble detected nerve fibres. **h** S100β. Software-based detection of 'IHC marker stain (brown chromogen)'-stained nerve fibres in AM-affected tissue. Thickened nerve fibres compared to healthy tissue. Red areas resemble detected positive nerve fibres. Digitalized using a digital slide scanner in bright-field mode (Pannoramic Flash 250 for bright-field, 3D-Histech AG, Budapest, Hungary) and processed by applying Definiens Developer Tissue Studios 4.3

specificity, positive predictive value and F1 score on the train and development sets and was finally evaluated on the test set. After 10-fold cross-validation, we achieved a final positive predictive value of 83.0%, sensitivity of 87.5%, specificity of 80% and F1 score of 88.9% on the test set. Evaluation metrics on the development set had a 95.0% positive predictive value, a sensitivity of 90.4%, a specificity of 85% and a F1 score of 92.3%. The negative predictive value was 100% in both cases. In the second approach in which we additionally split up the AM and non-AM groups into two subgroups consisting of suction biopsies and resections, it had no positive effect on the classifiers. It even slightly reduced sensitivity (data not shown).

Discussion

In this study, we applied pathology software, based on 'machine learning' with the objective to diagnose AM in submucosal immunohistochemically stained specimens. The highest sensitivity in diagnosis is still achieved by experienced pathologists employing H&E-stained specimens to detect neurons within ganglia (Kapur 2006). The mucosa and submucosa of the human large intestine contain several different nervous plexus with various types of neurons (Beuscher et al. 2014; Brehmer et al. 2010; Kramer et al. 2011). The submucosal plexus has been established as a diagnostic area for diagnosing AM (Bachmann et al. 2015).

By applying immunohistochemical markers, ganglia and nerve fibres in the submucosal tissue of the colon can be detected and analysed. The growing digitalization and automated analysis with software introduces new opportunities to process tissue slides and will certainly facilitate pathological diagnosis in the future (Madabhushi and Lee 2016). As diagnosis of Hirschsprung disease using immunohistochemistry is based on manual ganglia and nerve fibre detection, analysis and quantification within certain high-power fields (HPF), a software-based approach could enhance this process (Bachmann et al. 2015). The software is able to identify and quantify the entire submucosal tissue and does not focus on randomly selected HPF or oversee characteristic features.

Calretinin

Calretinin immunohistochemically stains most ganglia in the human colon and therefore enables the distinction between AM tissue and non-affected tissue. (Barshack et al. 2004). The role of calretinin as reference in cases of insufficient bioptic material, complete aganglionosis, inexperienced observers and in combination with acetylcholinesterase has already been stated by several studies (de Arruda Lourencao et al. 2013; Gonzalo and Plesec 2013; Guinard-Samuel et al. 2009). As there is essential variation between AM-positive and AM-negative samples, a lack of calretinin-reactive ganglion cells in the submucosa is distinctive for AM (Bachmann et al. 2015). In this study, calretinin-positive ganglion cells were detected and quantified by an algorithm, which has been trained on our pre-processed data with specific characteristics to minimize the rate of false negative detected cells and

Table 2 Representative presentation of measurements and statistics acquired for the features detected by our software approach for the different stainings. Differences between AM and non-AM tissues with no distinction between suction biopsies and resected material. Calretinin measures % area IHC marker stain; MAP2 measures % area IHC marker stain; S100 measures % area IHC marker stain; GLUT1 measures median vessel wall thickness

	Calretinin		MAP2		S100		GLUT1	
Diagnosis	AM	Non-AM	AM	Non-AM	AM	Non-AM	AM	Non-AM
Std	0.46%	1.26%	0.005%	0.08%	0.9%	0.8%	2.5 μm	1.5 μm
Mean	0.55%	1.44%	0.004%	0.07%	1.2%	0.9%	6.0 μm	5.7 μm
Min	0.17%	0.14%	0.0001%	0.001%	0.3%	0.2%	2.6 μm	3.9 μm
Max	1.59%	5.68%	0.02%	0.27%	3.6%	2.2%	9.6 μm	8.4 μm
95% confidence interval	0.15–0.94%	0.83–2.05%	0.0004–0.007%	0.03–0.1%	0.4–1.6%	0.8–1.6%	4.5–7.6 μm	4.7–6.6 μm
p value	0.012	0.0002	0.25	0.34				

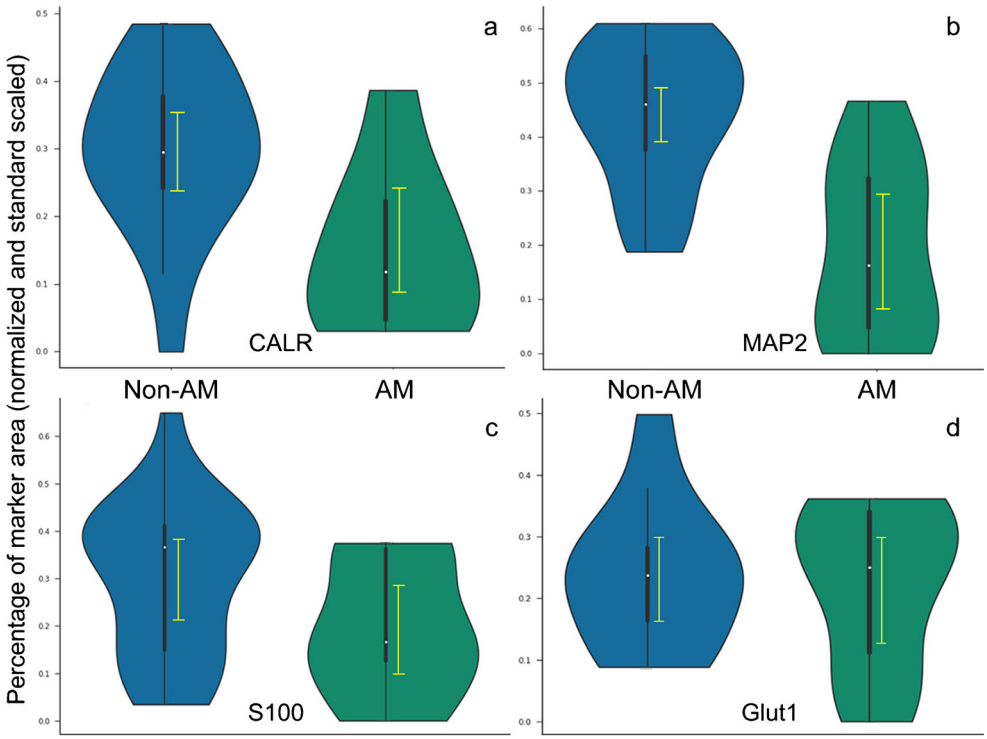


Fig. 3 Standard scaled (to uniform variance) and normed (to the range between 0 and 1) violin plots for the respective stainings for AM-affected and non-AM tissue samples. Within the violin plots, the 95% confidence intervals are depicted (yellow) and the boxplots are additionally shown in black for visualizing variance and the mean (white dots) of the data distribution

artefacts. The relation of the 'IHC marker stain (brown chromogen)'-marked area of ganglion cells to the total area of the submucosal tissue was taken as a measure for the quantity of neurons; however, the complete submucosa has been examined. The difference of marked area in relation to submucosal area was highly significant between AM-positive and non-AM tissues. This allowed a clear differentiation between AM-affected and not affected patients. However, the transition zone could not be evaluated here. Compared to MAP2, however, calretinin marks more ganglia and neurons but shows a significant quantity of false negative ganglia (Bachmann et al. 2015).

MAP2

MAP2 used for marking neuronal structures is, similar to calretinin, established as an alternative marker to diagnose AM (Caceres et al. 1984; Yang and Oh 2013). As a second criterion in inconclusive cases, MAP2 may substitute another required biopsy (Burtelow and Longacre 2009). Comparable to calretinin, MAP2-marked ganglion cells were automatically detected and quantified by our software with exclusion of false positive structures. The total area of marked ganglia was lower in MAP2-stained tissue than in calretinin samples. However, differences in relation of 'IHC marker stain (brown

Table 3 Representative presentation of measurements and statistics acquired for the features detected by our software approach for the different stainings. Differences between suction biopsies and resection material for the markers calretinin and MAP2. No significant differences could be found between suction biopsies and resection material comparing AM and non-AM tissues. Calretinin measures % area IHC marker stain; MAP2 measures % area IHC marker stain

	Calretinin				MAP2			
Diagnosis	AM (r)	AM (s)	Non-AM (r)	Non-AM (s)	AM (r)	AM (s)	Non-AM (r)	Non-AM (s)
Std	2.50%	1.25%	0.11%	0.73%	0.04%	0.1%	0.05%	0.008%
Mean	1.97%	1.44%	0.42%	0.82%	0.04%	0.08%	0.03%	0.007%
Min	0.48%	0.14%	0.29%	0.18%	0.01%	0.0002%	0.003%	0.0002%
Max	5.69%	2.68%	0.56%	1.47%	0.08%	0.27%	0.08%	0.018%
95% confidence interval	− 2.00–5.94%	0.79–1.71%	0.25–0.60%	− 0.34–1.98%	− 0.001 − 0.09%	0.02–0.13%	− 0.09–0.14%	− 0.001–0.016%
p value	0.35	0.33	0.47	0.26				

chromogen)' ganglia to submucosal tissue were highly significant between AM-positive and AM-negative samples. As stated above, MAP2 staining in our case did not allow the evaluation of the transition zone either.

S100β

Usually used as tumour and trauma marker, S100β identifies post-surgical insufficient innervation (Joosten et al. 1989) and marks glial cells (Holland et al. 2010). Bachmann et al. discovered a significant occurrence in AM-negative tissue with regard to diameter and amount of nerve fibres (Bachmann et al. 2015). Indeed, there is no significant coherence between S100β-marked nerve fibres in AM-positive tissue or transition zone, yet (Bachmann et al. 2015). S100β-stained nerve fibres were also detected and quantified by an 'IHC marker stain (brown chromogen)'-sensitive algorithm. The identified area resembled the size of nerve fibres of submucosal tissue. However, our study demonstrated no significant correlation between the size of nerve fibres and AM-affected tissue. Thus, digital detection and analysis of S100β staining were not helpful in AM diagnosis.

GLUT1

The protein GLUT1 has been identified as reliable marker for perineuria in AM specimens. Comparable to acetylcholinesterase staining, GLUT1 is capable of marking nerve fibres (Kakita et al. 2000). GLUT1 staining enables an evaluation of the size and quantity of nerve fibres. It turned out that there was a significant difference between AM-affected tissue displaying GLUT1-reactive perineurial sheaths and non-AM specimens, which was characterized by the absence of these positive perineurial sheaths (Bachmannet al. 2015; Kakita et al. 2000). However, this staining has just a considerable benefit for making an AM diagnosis if there is enough submucosal tissue available; otherwise, acetylcholinesterase was reported to be superior (Bachmann et al. 2015). We detected the perineuria of nerve fibres owing to the similarity of vessel walls and GLUT1-marked perineuria. Thus, the software made it possible to recognize and quantify latter, by using 'nerve fibre recognition and measurement function'. By configuring the algorithm to measure the vessel wall thickness, it was possible to evaluate the mean diameter of perineuria resembled by the median vessel wall thickness.

By configuring the algorithm to measure the thickness of nerve perineuria, it was possible to evaluate the mean diameter of perineuria. We did not find a distinct diagnostic relation between the data for AM-affected and healthy tissues. To be more precise, we conclude that thickness of perineuria might not be a needed feature for diagnostics of HP disease.

Conclusion of Results

This study demonstrates that computer-assisted analysis of calretinin and MAP2 stainings is sufficient to diagnose AM. Contrary to Bachmann et al. (Bachmann et al. 2015), S100β and GLUT1 did not have an additional benefit with regard to digital analysis. However, in uncertain cases, S100β and GLUT1 can be helpful to diagnose AM by an experienced pathologist. Additionally, no significant difference could be found between resection and suction biopsies for all markers applied which is in line with prevailing gold standard, the suction biopsy, as mentioned in the introduction. Suction biopsy is the gold standard in AM diagnosis and just in case of inconclusive histological diagnosis, a section biopsy is necessary (Andrassy et al. 1981; Dobbins and Bill 1965; Harjai 2000; Holland et al. 2010). However, by applying ordinary forceps biopsies, it is often not possible to obtain enough submucosal tissue; so for this reason they should be avoided (Koletzko et al. 1999; Wedel et al. 1999).

Digital pathology in combination with digital slide scanners is establishing various new opportunities to identify different types of tissue and epithelium as inflammation- or cancer-induced tissue aberrations (Irshad et al. 2014; Linder et al. 2012). By using 'machine and deep learning', it is indeed possible to train a 'computer pathologist' to recognize various structures, however, depending on the characteristics of latter. Currently, one limitation of completely automated pathology is the necessity of a manually adapted definition of certain areas within the digitized slides by a pathologist. To achieve a diagnostically conclusive result, it turns out to be necessary to combine both manual-adapted ROI detection and automated cellular analysis via 'deep learning'. Wang et al. described an advantage combining these options to exclude interference triggered by the mass of data or the missing sense for histological structures (Wang et al. 2014).

'Deep learning' depends on plenty of data to train a neural network algorithm (Madabhushi and Lee 2016). By defining training data from 27 patients, a certain pattern was recognizable, despite the comparatively small quantity of test persons. By increasing the number of slides/images, the algorithm is able to enhance the unsupervised cellular analysis as the software develops the ability to memorize disease-specific features and patterns (Madabhushi and Lee 2016).

The larger the available sample pool, the more detailed the abstraction capabilities of the software. Other than with the help of a supervised machine learning approach, which we used in this study, an unsupervised machine learning approach could also be taken into account. Supervised machine learning techniques are simply put an input to output mapping. An algorithm can be trained to predict distinct predefined parameters. Unsupervised machine learning techniques, however, are used to find yet unrecognized relationships.

By an increasing number of AM specimens, an unsupervised machine learning algorithm would potentially be able to recognize a significant relation between so far unknown parameters to diagnosis of AM or non-AM. However, to our knowledge, unsupervised machine learning techniques have not been used for diagnosis in histopathology up to this date (Deo 2015; Hastie et al. 2009).

While this study suggests that it is possible to diagnose AM by applying pathology software-based tissue analysis, the attempt to facilitate the diagnosis of AM by using detection and analysis software is not yet practicable for daily routine. The effort to train pathology software in order to recognize characteristic parameters of AM may be a simplification compared to manual diagnosis by experienced pathologists in the future. From a technical point of view, there are various challenges that will have to be overcome. Usually, the digitized images of slides show certain variability. As all stainings vary in their intensities and not every slide scanner produces the same resolution and contrast, the digital images differ. This becomes clear with regard to artefacts and colour intensity influenced by processing and staining the specimens. Especially the 'IHC marker stain (brown chromogenity)' varies in calretinin-, MAP2- and also S100β-stained images caused by different brightness of slide images after digitalization. In addition, artefacts can be mistaken as ganglia if several tissue layers overlap and create an 'IHC brown'-like colouration. Therefore, software-based detection and analysis features have issues providing identical data by examining digital images.

With regard to enhancing the efficiency of this digital analytic process, a greater pool of AM cases has to be investigated. As a side effect, the more cases and specimens are analysed also across different centres and their accompanying variations, the more robust the learned features are and hence the algorithm would be more reliable. However, AM is a rare disorder; it was not possible for us to this time point to provide a greater number of cases. Thus, a multicentre study would therefore improve the above-mentioned deficiencies and will be approached in future research projects and collaborations.

Image analysis based on software processing may be able to simplify the common way to diagnose AM only in the future. This idea is reasonable to be followed up if automated algorithms can be trained with less effort and less data.

Acknowledgements The present work was performed in fulfilment of the requirements of the Friedrich-Alexander Universität Erlangen-Nürnberg (FAU) for obtaining the degree 'Dr. med.dent' of Florian Schilling. We would like to thank Nicole Fuhrich and Definines Support for their excellent technical assistance.

Compliance with ethical standards

The use of the FFPE materials was approved by the local ethics committee (approval number: 85_12b, date 04/19/2012).

Conflict and interest statement The authors declare that they have no conflict of interests.

Informed consent Informed consent was obtained from all individual participants for whom identifying information is included in this article.

References

Aldridge RT, Campbell PE (1968) Ganglion cell distribution in the normal rectum and anal canal. A basis for the diagnosis of Hirschsprung's disease by anorectal biopsy. J Pediatr Surg 3:475–490

Andrassy RJ, Isaacs H, Weitzman JJ (1981) Rectal suction biopsy for the diagnosis of Hirschsprung's disease. Ann Surg 193:419–424

Bachmann L, Besendorfer M, Carbon R, Lux P, Agaimy A, Hartmann A, Rau TT (2015) Immunohistochemical panel for the diagnosis of Hirschsprung's disease using antibodies to MAP 2, calretinin, GLUT1 and S100. Histopathology 66:824–835

Badner JA, Sieber WK, Garver KL, Chakravarti A (1990) A genetic study of Hirschsprung disease. Am J Hum Genet 46:568–580

Barnes PR, Lennard-Jones JE, Hawley PR, Todd IP (1986) Hirschsprung's disease and idiopathic megacolon in adults and adolescents. Gut 27:534–541

Barshack I, Fridman E, Goldberg I, Chowers Y, Kopolovic J (2004) The loss of calretinin expression indicates aganglionosis in Hirschsprung's disease. J Clin Pathol 57:712–716

Bekkali N, Hamers SL, Schipperus MR, Reitsma JB, Valerio PG, Van Toledo L, Benninga MA (2008) Duration of meconium passage in preterm and term infants. Arch Dis Child Fetal Neonatal Ed 93: F376–F379

Beuscher N, Jabari S, Strehl J, Neuhuber W, Brehmer A (2014) What neurons hide behind calretinin immunoreactivity in the human gut? Histochem Cell Biol 141:393–405

Bodian M, Carter OO (1963) A family study of Hirschsprung's disease. Ann Hum Genet 26:261–277

Brehmer A, Rupprecht H, Neuhuber W (2010) Two submucosal nerve plexus in human intestines. Histochem Cell Biol 133:149–161

Burkardt DD, Graham JM Jr, Short SS, Frykman PK (2014) Advances in Hirschsprung disease genetics and treatment strategies: an update for the primary care pediatrician. Clin Pediatr (Phila) 53:71–81

Burtelow MA, Longacre TA (2009) Utility of microtubule associated protein-2 (MAP-2) immunohistochemistry for identification of ganglion cells in paraffin-embedded rectal suction biopsies. Am J Surg Pathol 33:1025–1030

Butler Tjaden NE, Trainor PA (2013) The developmental etiology and pathogenesis of Hirschsprung disease. Transl Res 162:1–15

Caceres A, Binder LI, Payne MR, Bender P, Rebhun L, Steward O (1984) Differential subcellular localization of tubulin and the microtubule-associated protein MAP 2 in brain tissue as revealed by immunocytochemistry with monoclonal hybridoma antibodies. J Neurosci 4: 394–410

Clark DA (1977) Times of first void and first stool in 500 newborns. Pediatrics 60:457–459

de Arruda Lourencao PL, Takegawa BK, Ortolan EV, Terra SA, Rodrigues MA (2013) A useful panel for the diagnosis of Hirschsprung disease in rectal biopsies: calretinin immunostaining and acetylcholinesterase histochesmistry. Ann Diagn Pathol 17: 352–356

Definiens A (2015) Definiens Tissue Studio 4.3 User Guide. Defineins Documentation München, Germany

Deo RC (2015) Machine learning in medicine. Circulation 132:1920–1930

Dobbins WO 3rd, Bill AH Jr (1965) Diagnosis of Hirschsprung's disease excluded by rectal suction biopsy. N Engl J Med 272:990–993

Duhamel B (1960) A new operation for the treatment of Hirschsprung's disease. Arch Dis Child 35:38–39

Gonzalo DH, Plesec T (2013) Hirschsprung disease and use of calretinin in inadequate rectal suction biopsies. Arch Pathol Lab Med 137: 1099–1102

Guinard-Samuel V, Bonnard A, De Lagausie P, Philippe-Chomette P, Alberti C, El Ghoneimi A, Peuchmaur M, Berrebi-Binczak D (2009) Calretinin immunohistochemistry: a simple and efficient tool to diagnose Hirschsprung disease. Mod Pathol 22:1379–1384

Harjai MM (2000) Hirschsprung's disease: revisited. J Postgrad Med 46: 52–54

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, Second Edition. NY Springer Science & Business Media New York

Holland SK, Hessler RB, Reid-Nicholson MD, Ramalingam P, Lee JR (2010) Utilization of peripherin and S-100 immunohistochemistry in the diagnosis of Hirschsprung disease. Mod Pathol 23:1173–1179

Irshad H, Veillard A, Roux L, Racoceanu D (2014) Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential. IEEE Rev Biomed Eng 7:97–114

Joosten K, Pruszczynski M, Severijnen RS, Festen C (1989) Causes of late complications in children operated on for Hirschsprung's disease: a preliminary immunohistochemical investigation using polyclonal antibodies against S-100 protein. Z Kinderchir 44:213–215

Kakita Y, Oshiro K, O'Briain DS, Puri P (2000) Selective demonstration of mural nerves in ganglionic and aganglionic colon by immunohistochemistry for glucose transporter-1: prominent extrinsic nerve pattern staining in Hirschsprung disease. Arch Pathol Lab Med 124: 1314–1319

Kapur RP (2006) Can we stop looking? Immunohistochemistry and the diagnosis of Hirschsprung disease. Am J Clin Pathol 126:9–12

Kapur RP (2014) Calretinin-immunoreactive mucosal innervation in very short-segment Hirschsprung disease: a potentially misleading observation. Pediatr Dev Pathol 17:28–35

Kapur RP, Kennedy AJ (2012) Transitional zone pull through: surgical pathology considerations. Semin Pediatr Surg 21:291–301

Kapur RP, Reed RC, Finn LS, Patterson K, Johanson J, Rutledge JC (2009) Calretinin immunohistochemistry versus acetylcholinesterase histochemistry in the evaluation of suction rectal biopsies for Hirschsprung disease. Pediatr Dev Pathol 12:6–15

Koletzko S, Jesch I, Faus-Kebetaler T, Briner J, Meier-Ruge W, Muntefering H, Coerdt W, Wessel L, Keller KM, Nutzenadel W, Schmittenbecher P, Holschneider A, Sacher P (1999) Rectal biopsy for diagnosis of intestinal neuronal dysplasia in children: a prospective multicentre study on interobserver variation and clinical outcome. Gut 44:853–861

Kramer K, da Silveira AB, Jabari S, Kressel M, Raab M, Brehmer A (2011) Quantitative evaluation of neurons in the mucosal plexus of adult human intestines. Histochem Cell Biol 136:1–9

Linder N, Konsti J, Turkki R, Rahtu E, Lundin M, Nordling S, Haglund C, Ahonen T, Pietikainen M, Lundin J (2012) Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. Diagn Pathol 7:22

Lopez-Alonso M, Ribas J, Hernandez A, Anguita FA, Gomez de Terreros I, Martinez-Caro A (1995) Efficiency of the anorectal manometry for the diagnosis of Hirschsprung's disease in the newborn period. Eur J Pediatr Surg 5:160–163

Madabhushi A, Lee G (2016) Image analysis and machine learning in digital pathology: challenges and opportunities. Med Image Anal 33:170–175

Martin LW, Altemeier WA (1962) Clinical experience with a new operation (modified Duhamel procedure) for Hirschsprung's disease. Ann Surg 156:678–681

Martucciello G (2008) Hirschsprung's disease, one of the most difficult diagnoses in pediatric surgery: a review of the problems from clinical practice to the bench. Eur J Pediatr Surg 18:140–149

Meier-Ruge W, Lutterbeck PM, Herzog B, Morger R, Moser R, Scharli A (1972) Acetylcholinesterase activity in suction biopsies of the rectum in the diagnosis of Hirschsprung's disease. J Pediatr Surg 7:11–17

Murphy F, Puri P (2005) New insights into the pathogenesis of Hirschsprung's associated enterocolitis. Pediatr Surg Int 21:773–779

Pratap A, Gupta DK, ea TA (2007) Applicationofaplainab- dominal radiograph transition zone (PARTZ) in Hirschsprung's disease. BMC Pediatr 7:5

Qualman SJ, Jaffe R, Bove KE, Monforte-Munoz H (1999) Diagnosis of hirschsprung disease using the rectal biopsy: multi-institutional survey. Pediatr Dev Pathol 2:588–596

Sarnat HB, Seagram GF, Trevenen CL, Rubin SZ (1985) A fluorochromic stain for nucleic acids to demonstrate submucosal and myenteric neurons in Hirschsprung's disease. Am J Clin Pathol 83:722–725

Szylberg L, Marszalek A (2014) Diagnosis of Hirschsprung's disease with particular emphasis on histopathology. A systematic review of current literature. Prz Gastroenterol 9:264–269

Taguchi T, Tanaka K, Ikeda K (1985) Immunohistochemical study of neuron specific enolase and S-100 protein in Hirschsprung's disease. Virchows Arch A Pathol Anat Histopathol 405:399–409

Volpe A, Alaggio R, Midrio P, Iaria L, Gamba P (2013) Calretinin, beta-tubulin immunohistochemistry, and submucosal nerve trunks morphology in Hirschsprung disease: possible applications in clinical practice. J Pediatr Gastroenterol Nutr 57:780–787

Wang H, Cruz-Roa A, Basavanhally A, Gilmore H, Shih N, Feldman M, Tomaszewski J, Gonzalez F, Madabhushi A (2014) Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. J Med Imaging (Bellingham) 1:034003

Wedel T, Roblick U, Gleiss J, Schiedeck T, Bruch HP, Kuhnel W, Krammer HJ (1999) Organization of the enteric nervous system in the human colon demonstrated by wholemount immunohistochemistry with special reference to the submucous plexus. Ann Anat 181: 327–337

Yang WI, Oh JT (2013) Calretinin and microtubule-associated protein-2 (MAP-2) immunohistochemistry in the diagnosis of Hirschsprung's disease. J Pediatr Surg 48:2112–2117

Yanlei H, Anupama B, Shan Z, Xianmin X, Lian C (2011) The expression of enteric nerve markers and nerve innervation in total colonic aganglionosis. Int J Surg Pathol 19:303–308