



A deep learning-based histopathology classifier for Focal Cortical Dysplasia

Jörg Vorndran^{1,20} · Christoph Neuner^{1,20} · Roland Coras^{1,20} · Lucas Hoffmann^{1,20} · Simon Geffers¹ · Jonas Honke¹ · Jochen Herms² · Sigrun Roeber² · Hajo Hamer^{3,20} · Sebastian Brandner^{4,20} · Till Hartlieb^{5,6,20} · Tom Pieper⁶ · Manfred Kudernatsch^{6,7} · Christian G. Bien⁸ · Thilo Kalbhenn⁹ · Matthias Simon⁹ · Homa Adle-Biasette^{10,11} · Jesús Cienfuegos¹² · Roberta Di Giacomo^{13,20} · Rita Garbelli^{13,20} · Hajime Miyata¹⁴ · Angelika Mühlebner^{15,20} · Savo Raicevic¹⁶ · Tuomas Rauramaa^{17,20} · Fabio Rogerio^{18,19} · Ingmar Blümcke^{1,20} · Samir Jabari^{1,20}

Received: 11 August 2022 / Accepted: 14 February 2023 / Published online: 8 March 2023
© The Author(s) 2023

Abstract

A light microscopy-based histopathology diagnosis of human brain specimens obtained from epilepsy surgery remains the gold standard to confirm the underlying cause of a patient's focal epilepsy and further inform postsurgical patient management. The differential diagnosis of neocortical specimens in the realm of epilepsy surgery remains, however, challenging. Herein, we developed an open access, deep learning-based classifier to histopathologically assess whole slide microscopy images (WSI) and to automatically recognize various subtypes of Focal Cortical Dysplasia (FCD), according to the ILAE consensus classification update of 2022. We trained a convolutional neuronal network (CNN) with fully digitalized WSI of hematoxylin–eosin stainings obtained from 125 patients covering the spectrum of mild malformation of cortical development (mMCD), mMCD with oligodendroglial hyperplasia in epilepsy (MOGHE), FCD ILAE Type 1a, 2a and 2b using 414 formalin-fixed and paraffin-embedded archival tissue blocks. An additional series of 198 postmortem tissue blocks from 59 patients without neurological disorders served as control to train the CNN for homotypic frontal, temporal and occipital areas and heterotypic Brodmann areas 4 and 17, entorhinal cortex and dentate gyrus. Special stains and immunohistochemical reactions were used to comprehensively annotate the region of interest. We then programmed a novel tile extraction pipeline and graphical dashboard to visualize all areas on the WSI recognized by the CNN. Our deep learning-based classifier is able to compute $1000 \times 1000 \mu\text{m}$ large tiles and recognizes 25 anatomical regions and FCD categories with an accuracy of 98.8% (F1 score = 0.82). Microscopic review of regions predicted by the network confirmed these results. This deep learning-based classifier will be made available as online web application to support the differential histopathology diagnosis in neocortical human brain specimens obtained from epilepsy surgery. It will also serve as blueprint to build a digital histopathology slide suite addressing all major brain diseases encountered in patients with surgically amenable focal epilepsy.

Keywords Cortex · Epilepsy · Digital pathology · Deep learning · Classification · FCD2 · MOGHE · mMCD · FCD1 · Convolutional neuronal network

1 Introduction

Advancements in WSI technology helped to further promote digital pathology and the development of artificial intelligence (AI)-based disease classification algorithms at a larger scale [8, 14]. However, published AI studies cover only few topics of the diagnostic spectrum in neuro- or

histopathology including epilepsy surgery, yet. Recently, a deep learning-based algorithm was able to demonstrate the ability to differentiate the cellular profile of FCD Type 2b from cortical tuber on routine hematoxylin–eosin (HE)-stained WSI [19]. Their algorithm recognized bulky and strand-like matrix reaction, halo-like balloon cell artifacts as well as bigger astroglial cell nuclei in cortical tuber compared to FCD ILAE Type 2b as disease-specific signatures. These small differences at the cellular level and in

Extended author information available on the last page of the article

the texture were not recognized by light microscopy before and also helped to define a histopathology scoring system when training colleagues not specialized in this diagnostic arena. Another hurdle in the development of AI-based histopathology classifiers is the huge work burden for annotation of large datasets and which support the use of weakly supervised deep learning pipelines [8, 14]. This applies also for the prediction of defined genetic mutations in histopathology specimens, e.g., colorectal, gastric and bladder cancer [14]. A different AI-based approach was able to replicate histochemical and immunohistochemical stainings from unstained slides without actually performing the staining procedures [12, 29]. In contrast, AI-based classifiers are increasingly developed in neuroradiology in order to reduce the burden of reviewing large stacks of MR images at various protocols and to detect small lesions in drug-resistant focal epilepsy, i.e., FCD [15]. However, the full potential of cost-effective and AI-based digital pathology classifiers needs to be further explored.

Malformations of cortical development (MCD) represent the first most common structural brain lesion in children with drug-resistant focal epilepsy. Focal Cortical Dysplasia (FCD) represents approx. 75% of all MCD entities collected in large epilepsy surgery series [6, 20]. The International League against Epilepsy (ILAE) has proposed a consensus classification scheme for FCD in order to cover the clinical spectrum of FCD, their associated histopathology patterns and likely etiology, e.g., brain somatic mutations in the mTOR pathway [3, 4, 7]. The histopathological classification of these categories remains, however, an ever-challenging issue in daily routine practice [22, 23], and which has been documented many times in the scientific literature [1, 4, 5, 9, 11]. Molecular neuropathology has been recognized as a helpful diagnostic tool to objectively classify human brain lesions, e.g., in brain tumors, and was recently introduced also into the realm of epileptology and epilepsy surgery, e.g., FCD ILAE Type 2 with brain somatic mTOR pathway mosaicism or MOGHE, SLC35A2 altered [3, 4]. These entities have been also included into the ILAE consensus FCD classification update of 2022, which mandates an integration of clinical, histopathological and neuroimaging information into the final FCD diagnosis [22]. Molecular testing of surgical human brain samples remains, however, cost intensive and may not be accessible to many surgical centers and pathology laboratories around the world. An AI-based histopathology classifier for FCD would be cost-effective, available online and thus improve the knowledge and access to routine histopathology diagnosis. In the current work, we retrieved a large series of MCD specimens covering the spectrum of common and rare FCD subtypes, used the WSI technology and developed a supervised, deep learning-based histopathology classifier to

reliably detect FCD subtypes in surgically resected human brain tissue specimens.

2 Materials and methods

2.1 Patients included in this study

Archival microscopy slides of tissue specimens from 125 patients with the histopathological diagnosis of mMCD, MOGHE, FCD Type 1a, 2a and 2b according to the ILAE classification scheme [7, 22] were retrieved from the Neuropathological Institute at Universitätsklinikum Erlangen (see supplemental Table 1 and supplemental Figure 1). In addition, we retrieved age-matched postmortem brain tissue of Brodmann areas 4 and 17, entorhinal cortex and dentate gyrus from 59 patients without any neurological disorder from the Neuropathological Institute at Universitätsklinikum Erlangen and the Ludwig-Maximilian University München. Histological sections were 3–5 µm thin, obtained from formalin-fixed and paraffin-embedded (FFPE) tissue blocks and stained with hematoxylin–eosin (HE) and HE–Luxol–Fast–Blue (HE–LFB) according to routine neuropathology workup protocols [2]. Additional HE, HE–LFB or immunohistochemical stainings for NeuN, MAP2, SMI32, and Olig2 epitopes were prepared from archival FFPE tissue blocks when necessary using the same protocols. All slides were microscopically reviewed by experienced neuropathologists (IB, JoHe, SR and RC) to select a minimum of 20 cases from each histopathology entity to be included in this study (see supplemental Table 1). The FCD1a category included those patients described recently [17]. The glossary summarizes the definition applied for each category. Whole slide images (WSI) were digitally recorded from HE as well as adjacent immunohistochemical stainings using a Hamamatsu S60 scanner (Hamamatsu Photonics Europe, Herrsching, Germany) with a mean of 3.02 Gb per slide, ranging from 0.47 to 5.72 Gb. Each patient's clinical history was retrieved from the archival hospital files. The study was approved by the University of Erlangen ethical review board under the agreement number 193_18B.

Glossary of 25 anatomo-pathological categories

BA04-PM	Postmortem FFPE tissue sample characterized by heterotypic neocortical architecture of Brodmann area 4 with giant pyramidal neurons of Betz. Immunohistochemistry for SMI32—neurofilament staining was made available for annotations
BA17-PM	

	Postmortem FFPE tissue sample characterized by heterotypic neocortical architecture of Brodmann area 17 including the strip of Gennari. HE-LFB staining was made available for annotation.		directed against non-phosphorylated neurofilament epitopes (SMI)
OCX-PM	Postmortem FFPE tissue sample characterized by homotypic neocortical architecture, adjacent to Brodmann area 17. HE-LFB staining was made available for annotation	FCD2a-hCx	Surgical FFPE tissue obtained from the homotypic neocortex adjacent to the FCD2a lesion
OCX-SX	Surgical FFPE tissue samples characterized by homotypic neocortical architecture, obtained from occipital lobe resections in patients included in the FCD1a category (see therein)	FCD2a-WM	Surgical FFPE tissue of the white matter adjacent to the FCD2a lesion
FLX-PM	Postmortem FFPE tissue characterized by homotypic neocortical architecture of the frontal lobe	FCD2b	Surgical FFPE tissue characterized by abnormal cortical architecture, dysmorphic neurons and balloon cells as evidenced by immunohistochemistry for SMI32 and vimentin
TLX-PM	Postmortem FFPE tissue characterized by homotypic neocortical architecture of the temporal lobe, obtained from the midlevel of the hippocampal body	FCD2b-hCx	Surgical FFPE tissue characterized by homotypic neocortical architecture adjacent to the FCD2b lesion
ECX-PM	Postmortem FFPE tissue characterized by the heterotypic entorhinal cortex architecture. No further immunohistochemistry was performed	FCD2b-WM	Surgical FFPE tissue characterized by the subcortical white matter adjacent to FCD2b lesion
DG-PM	Postmortem FFPE tissue characterized by the granule cell layer of the Dentate Gyrus and adjacent to the polymorphic molecular cell layers at the level of the mid-hippocampal body	NTU	“No tissue” describes the translucent WSI space around the stained tissue section
WM-PM	Postmortem FFPE tissue characterized by myelin containing white matter from occipital, frontal, temporal Lobe specimens. No further immunohistochemistry was performed	SUB-SX	Surgical FFPE tissue including the subarachnoidal space with their blood vasculature and meninges
WM-SX	Surgical FFPE tissue from the white matter of frontal, parietal, temporal and occipital lobes without any further evidence of histopathologic lesions as evidenced by a systematic immunohistochemical review using the epilepsy workup staining panel [2]	SUB-PM	Postmortem FFPE tissue characterized by the subarachnoidal space filled with blood vessels and the pia mater. No further immunohistochemical stainings were performed
mMCD	Surgical FFPE tissue characterized by subcortical white matter with excessive heterotopic neurons as histochemically confirmed with MAP2 antibodies. Tissue sections were obtained from frontal and temporal lobe regions. These areas should not include the crown of gyrus	The nine major categories were labeled in bold (see also supplemental Figure 1)	
mMCD-WM	Adjacent subcortical region of mMCD specimens without excessive heterotopic neurons as histochemically confirmed with MAP2 antibodies	2.2 Hardware and software configuration to compute the FCD classifier	
mMCD-hCx	Adjacent cortical region of mMCD specimens. We have not performed additional immunohistochemical stainings	We implemented the data extraction, data preparation, storage and the required arithmetic operations on a local computer unit with AMD CPU (AMD RYZEN THREA-DRIPPER 1900X 3.8 GHz 8-Core) and 64 Gb of RAM, running on Ubuntu 20.04.03 LTS. The tensor calculations were performed on a Gigabyte GeForce 2070 Super 8 GB, Cuda Version 11.2. The input data for the model calculation were stored and provided via SSD hard drives with a total storage capacity of 8 TB.	
MOGHE	Surgical FFPE tissue characterized by patchy oligodendroglial hyperplasia of the white matter preferably at the gray–white matter boundary as immunohistochemically confirmed with antibodies directed against OLIG2	We selected the ResNet18 CCN architecture from a pre-test of CNN architectures of the ResNet and VGG family, e.g., ResNet18, ResNet34 and VGG16. Larger networks of the ResNet family could not be tested due to our limited GPU resources. The ResNet18 CNN architecture (http://dx.doi.org/10.1109/CVPR.2016.90 , last visited 21.06.2022) was selected due to its fast and GPU-resource saving performance on the ImageNet data set (https://doi.org/10.1109/CVPR.2009.5206848) (data not shown) and was embedded in the open-source Python package and AI-project fastai (https://doi.org/10.3390/info11020108 , last visited 10.06.2022). Tensor-based calculations were based on the pytorch backend (https://doi.org/10.48550/arXiv.1912.01703 , last visited 15.06.2022). The basic network architecture was not changed with the exception of the	
MOGHE-hCx	Surgical FFPE tissue of the homotypic neocortex adjacent to the MOGHE		
FCD1a	Surgical FFPE tissue obtained from temporal, parietal and occipital lobe resections and histopathologically characterized by excessive neuronal microcolumns and immunohistochemically confirmed with antibodies directed against the NeuN epitope [17]		
FCD2a	Surgical FFPE tissue characterized by abnormal neocortical architecture and dysmorphic neurons as immunohistochemically evidenced with antibodies		

application-related classifier head in order to compute all defined tissue categories (see glossary). All data management of WSI tiles was performed on an open access library as recently published [24] and whose structure precisely reflects the required workflow of data generation, data management and data evaluation.

2.3 The annotation and extraction pipeline

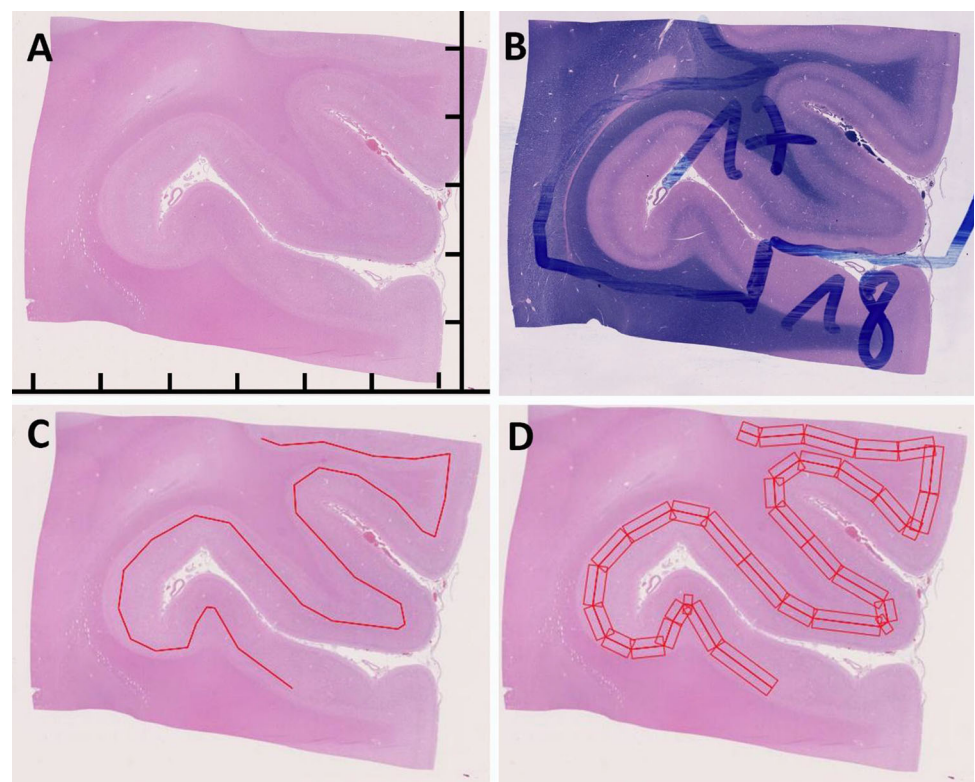
The region of interest (ROI) representing the anticipated anatomical region or cortical malformation was manually annotated on the microscopy slide or WSI using the Hamamatsu NDPi viewer by our experienced neuropathologists IB, RC, JHe, SR using HE-LFB and/or respective immunostainings (Fig. 1b). Regions not cut perpendicular to the cortical surface were excluded from the study. Annotations were digitally introduced onto the WSI of the adjacent HE staining (Fig. 1c). A Python script was developed to identify the vector coordinates of these annotations and to automatically extract 1mm^2 tiles with a stride of $100\text{ }\mu\text{m}$ at Hamamatsu ndpi-Level 2, perpendicular to the annotation line by using the open access OpenSlide (<https://doi.org/10.4103/2153-3539.119005>) and diagonal-crop libraries (GitHub—jobevers/diagonal-crop: Diagonally crop an image using python and pillow) (Fig. 1d). Tiles were then resampled at a resolution of 1024×1024 pixels and stored locally for later use by the

CNN adding the histopathology label to each unique filename.

2.4 Dataset splitting

In order to monitor and control the training process, we divided the total set of available training images into a training group and a validation group. To do this, we determined the total number of training images for each of our eight main entities to be examined (see Supplement Figure 1) and assigned 20% from each group for the validation process. The splitting process was carried out using a random number generator at the patient level and subsequently using an iteration selection, also at the patient level, until the required number of validation images had been separated, i.e., all images of a specific patient were available in either the training group or in the validation group. The aim of this approach was to prevent the system from assignments based on specific individual characteristics of the individual patient samples in order to positively support the generalization of the model and avoid overfitting in the context of hidden variables within the patient data, e.g., HE coloring patterns of a given case or certain repetitive artifacts.

Fig. 1 Annotation pipeline developed for the FCD classifier. **a** WSI of a postmortem FFPE sample from the occipital lobe stained with HE. The distance between each scale is $500\text{ }\mu\text{m}$. **b** For each category we used a special stain or immunostaining to reproducibly recognize anatomical landmarks or the histopathological lesion. Here, we delineated Gennari's strip of Brodmann area 17 on the microscopy glass slide with HE-LFB. **c** The same region was then annotated with npdi viewer on the WSI (same shown in **a**). **d** A Python script was developed to generate and extract tiles perpendicular to the annotation line



2.5 Metric output of the ResNet18 CNN

The prediction score given by the CNN is a measure of the likely probability of a tile to match with any of the introduced categories, i.e., “maximum confidence.” To achieve this value, we applied a sigmoid activation function for multi-label classification tasks to the array of values (digits) that was given by the final linear layer of a neuronal network for each image.

All metrics were obtained from the *fastai* library and adapted to our multi-label classification approach. *Fastai* itself hereby uses scikit-learn metrics and pytorch loss-functions. We monitored the training performance by `train_loss` and `validation_loss` as calculated by the `BCEWithLogitsLoss` function (threshold 0.5). Generally both parameters should converge toward zero as the number of epochs is rising. We determined overfitting of the training process as the divergence between training loss, i.e., numbers decrease, compared to validation loss, i.e., numbers increase, and other metrics, i.e., numbers decrease. `Train_loss` and `validation_loss` were therefore monitored to optimize computation time for training and to find the best time point to stop training the model.

In addition, we monitored `accuracy_multi`, `PrecisionMulti`, `RecallMulti` and `F1ScoreMulti` to evaluate the practicability of the trained models as well as overfitting. The `accuracy_multi` score was calculated from the logits score provided by the ResNet18 for each image. These logits were normalized therefore into a value range between 0 and 1 by using the sigmoid function. From these values, we defined a positive prediction at a threshold higher or equal 0.5 (Table 1). Values below this threshold were considered to not meet the prediction (negative).

These predictions were converted into a one-hot-encoded format and compared with the basic ground truth, i.e., the manually labeled values for the corresponding image. The mean of all element-wise compared values then provides the `accuracy_multi` score metric (see below).

We applied weighted averaging for `PrecisionMulti`, `RecallMulti` and `F1ScoreMulti` across all labels in this work to take the presence of unbalanced categories into account (see dataset splitting) [26]. This means that the calculated metrics for the respective individual classes are weighted with the proportion of training images in this class in relation to the size of the total population of training image data. Other averaging regimes were not used herein.

3 Results

A total of 612 WSI slides were obtained from 184 patients (see supplement Table 1), representing at least 20 independent patients for each of the main anatomic-pathological categories. With this dataset, we generated a total 533.461 tiles for further use in this study, labeled for 25 subcategories, including nine major histopathology categories of diagnostic interest and their helper classes (see glossary and supplemental Figure 1). Training of the ResNet18 CNN was carried out with a batch size of 16 tiles each and the Adams optimizer (<https://doi.org/10.48550/arXiv.1412.6980>). The learning process used the cyclic learning rate policy (<https://doi.org/10.1109/WACV.2017.58>). The maximum learning rate multiplier variable was set to 0.015, as determined by the *fastai* learning rate finder. The basic model was trained for 10 epochs (supplement Table 2, column 1). In addition, a fivefold cross-validation

Table 1 Example for calculation of `accuracy_multi`

Logit	Sigmoid	One-hot-encoded prediction after applied threshold	One-hot-encoded ground truth	Comparison	Label (example)
0.50	0.62	1	1	1	BA04-PM
1.10	0.75	1	1	1	BA17-PM
− 0.60	0.35	0	0	1	DG-PM
0.50	0.62	1	0	0	ECX-PM
− 1.20	0.23	0	0	1	FCD1a
1.30	0.79	1	0	0	FCD2a
...	...	...	...	...	...

1st column: logit values represent the not yet normalized output of prediction probabilities of the classifier for each WSI tile. 2nd column: The sigmoid function was applied to each logit to normalize the probability values into a value range between 0 = very unlikely and 1 = very likely. 3rd column: A threshold value was set to 0.5 to derive one-hot-encoded predictions from the probabilities shown in 2nd column. If the value in column 2 was above the threshold value (0.5), a positive prediction (1) is assigned for the presence of the category (= label in 6th column). Values below 0.5 were assigned as negative prediction (0). The real presence of one or more histopathology categories in the original WSI was reported as One-Hot coding ground truth in column 4. 5th column: The comparison of the model's prediction with the ground truth provides logical truth values for each label trained in the classifier. The summation of calculated truth values and their subsequent averaging resulted in the `accuracy_multi`-score for the input image

was carried out on the base of this model in which validation sets were unique across all fivefold (Table 2). Other sub-models were trained for 5 epochs depending on the achieved metrics score until no significant improvement in control parameters could be recognized (see supplement Table 2). The pre-trained convolution base was not changed during training for all models. All models mentioned above were re-initialized with the ImageNet weights before training. Data augmentation was applied during the training at batch level including horizontal and vertical mirroring, rotation up to 180°, image zoom from factor 0.9 to 1.1, and “zeros” as padding mode (*fastai*).

The training of the model was performed with a case cohort representing approx. 80% of each of the 25 categories was validated with the remaining cases, i.e., dataset splitting at the patient level. The ResNet18 CNN was fed with tiles from our comprehensive extraction routine and managed by the WSI processing pipeline [24]. The first run achieved an accuracy_multi of 0.986, however, which prompted us to stop the training after the tenth run at an accuracy_multi of 98.8% (F1MultiScore = 82.2%) within the best epoch (epoch nine, total computing time of approx. 16 h, supplement Table 2). Training progress ceased after the seventh training epoch. Any further training epoch did not show notable improvement in the control metrics (data not shown). However, we did not observe overfitting, i.e., discrepancies between train_loss and validation_loss, accuracy_multi and F1MultiScore at higher epochs (supplemental Figure 2). The advantage of dataset splitting at the patient level vs. random dataset splitting became further evident when challenging the two approaches with entirely new cases of genetically confirmed diagnosis never presented to the CNN (supplement Table 1). This analysis revealed a poorer generalization of the random-split model (supplement Table 3) and was not considered further, therefore.

Our basic model (Table 3) included all categories specified (see glossary) and achieved a maximum accuracy_multi score of 0.988 after the tenth epoch. The RecallMulti score for this basic model achieved 0.780 and showed more variation across our categories (Table 3). Best PrecisionMulti score and F1MultiScore with applied weighted averaging were calculated with 0.900 and 0.820, respectively. The accuracy_multi score was equally high

across the major anatomo-pathological categories (Table 3). Within this group, best accuracy_multi values were achieved for FCD2a (0.991) and least for mMCD (0.955). Overall, the highest values were achieved for DG-PM (1.000), along with WM-PM (1.000) followed by BA17-PM (0.998). WM-PM also achieved a high recall value (1.000), whereas recall for WM-SX reached 0.720. Samples of white matter (WM) adjacent to the FCD2a and FCD2b areas showed lower values for precision and recall (FCD2a-WM (0.290, 0.580), FCD2b-WM (0.450, 0.360). WM adjacent to mMCD could not be recognized by the model as a separate class. However, the evaluation results showed sufficient differentiation between cortex and white matter in general as did anatomical border zones of the subarachnoidal space (SUB) or the unstained glass slide with non-tissue background (NTU). Out of the 8 main categories, F1 score and recall were lowest for MOGHE (0.450/0.420), followed by mMCD (0.600, 0.460). FCD2b (0.830, 0.770) showed higher metrics compared to FCD2a (0.740, 0.670). Best results could be achieved for BA04-PM, BA17-PM and DG-PM (see Table 3).

3.1 A weakly supervised classifier approach did not achieve successful results

In another approach, we collected all slides from FCD1a, FCD2a, FCD2b, MOGHE and mMCD, extracted approx. 1×10^6 tiles from the entire tissue region of the WSI using the previously published library [24]. No specific annotation was applied in this approach, i.e., weakly supervised. All tiles were assigned to one of the five labels mentioned above. The ResNet18 CNN was trained with five epochs for approximately two weeks. Neither training nor validation accuracy reached a level above 50%. This experiment confirmed the importance and value of our supervised approach of time-costly histopathology annotation.

3.2 Calculating the most efficient number of labels/categories for the FCD classifier

Another question pertinent to our approach was to assess the best and most efficient number of discriminating categories or the combinations thereof, e.g., grouping together homotypic neocortices from postmortem and surgical

Table 2 Performance overview for fivefold cross-validation

Fold	1	2	3	4	5	Mean	SD
Accuracy_multi	0.988	0.986	0.988	0.984	0.988	0.987	0.001
PrecisionMulti	0.892	0.898	0.872	0.827	0.896	0.877	0.027
RecallMulti	0.778	0.764	0.796	0.747	0.783	0.774	0.017
F1ScoreMulti	0.818	0.800	0.818	0.764	0.822	0.804	0.022

Metrics for each cross-validation fold with overall mean (mean) and standard deviation (SD)

Table 3 Detailed metrics for labeled subcategories within the basic model

	Label	accuracy_multi	PrecisionMulti	RecallMulti	F1ScoreMulti
1	BA04-PM	0.9947	0.9500	0.9300	0.9400
2	BA17-PM	0.9981	0.9900	0.9500	0.9700
3	DG-PM	1.0000	1.0000	1.0000	1.0000
4	ECX-PM	0.9955	0.9800	0.4300	0.6000
5	FCD1a	0.9688	0.9400	0.5800	0.7200
6	FCD2a	0.9908	0.8200	0.6700	0.7400
7	FCD2a-WM	0.998	0.290	0.580	0.390
8	FCD2a-hCx	0.996	0.750	0.610	0.670
9	FCD2b	0.989	0.890	0.770	0.830
10	FCD2b-WM	0.984	0.450	0.360	0.400
11	FCD2b-hCx	0.972	0.970	0.430	0.590
12	FLX-PM	0.989	0.820	0.740	0.780
13	MOGHE	0.970	0.490	0.420	0.450
14	MOGHE-hCx	0.970	0.730	0.310	0.430
15	NTU	0.995	0.970	1.000	0.980
16	OCX-PM	0.994	0.680	0.650	0.660
17	OCX-SX	0.979	0.800	0.250	0.380
18	SUB-PM	0.998	0.980	0.950	0.970
19	SUB-SX	0.997	0.990	0.980	0.990
20	TLX-PM	0.991	0.940	0.880	0.910
21	WM-PM	1.000	1.000	1.000	1.000
22	WM-SX	0.963	0.760	0.720	0.740
23	mMCD	0.955	0.870	0.460	0.600
24	mMCD-WM	0.993	0.000	0.000	0.000
25	mMCD-hCx	0.989	0.900	0.590	0.710
Weighted average		0.988	0.900	0.780	0.820

label = BA04 Brodmann area 4; BA17 Brodmann area 17; PM post mortem; OCX occipital cortex; SX surgical sample; FLX frontal cortex; TLX temporal cortex; ECX entorhinal cortex; DG Dentate Gyrus; WM white matter; mMCD mild malformation of cortical development; hCx homotypic cortex; MOGHE mMCD with oligodendroglial hyperplasia; FCD1a Focal Cortical Dysplasia ILAE Type 1a; FCD2a Focal Cortical Dysplasia ILAE Type 2a; FCD2b Focal Cortical Dysplasia ILAE Type 2b; NTU no tissue; SUB subarachnoidal space; all defining histological features were specified in the Glossary; accuracy_multi, PrecisionMulti, RecallMulti, F1ScoreMulti were calculated as described in Material and Methods

tissues and of frontal, temporal and occipital regions, or grouping anatomical compartments of each disease category, i.e., white matter (WM) and adjacent cortices (hCx). The best performance of the CNN was achieved when all anatomo-pathological labels were included into the training and validation process (supplement Table 2). Each of the aforementioned groupings reached less accuracy. Along these lines, exclusion of categories such as all postmortem controls, no tissue regions (NTU) or adjacent FCD compartments did not achieve a better performance of the algorithm. In a next step, we developed a data class to manage calculated predictions and their graphical representation on the WSI by the trained model and to provide an interpretation aid, i.e., the histopathology classifier (Fig. 2). All digitally available information was extracted from the WSI by OpenSlide including the magnification

scale and spatial x and y dimension of each WSI image under study. This information was necessary to extract tiles with the same specification used to train the CNN. Then, the program was able to generate a grid of tiles covering the entire space of the WSI ($1000 \times 1000 \mu\text{m}$ and 1024×1024 pixels each; see Fig. 2a). The tiles were fed into the trained model in a png-format, and their prediction score was calculated (Fig. 2b). The results of this calculation were stored with reference to the respective patches. A raster image was created from these predicted values in the same dimension as the original patch raster. An algorithm then determined the average prediction probability for all categories recognized on the slide. Using an adjustable threshold value, categories with low predictive values were hidden, e.g., < 0.5 (Fig. 2b). A color and transparency value was calculated for categories above the

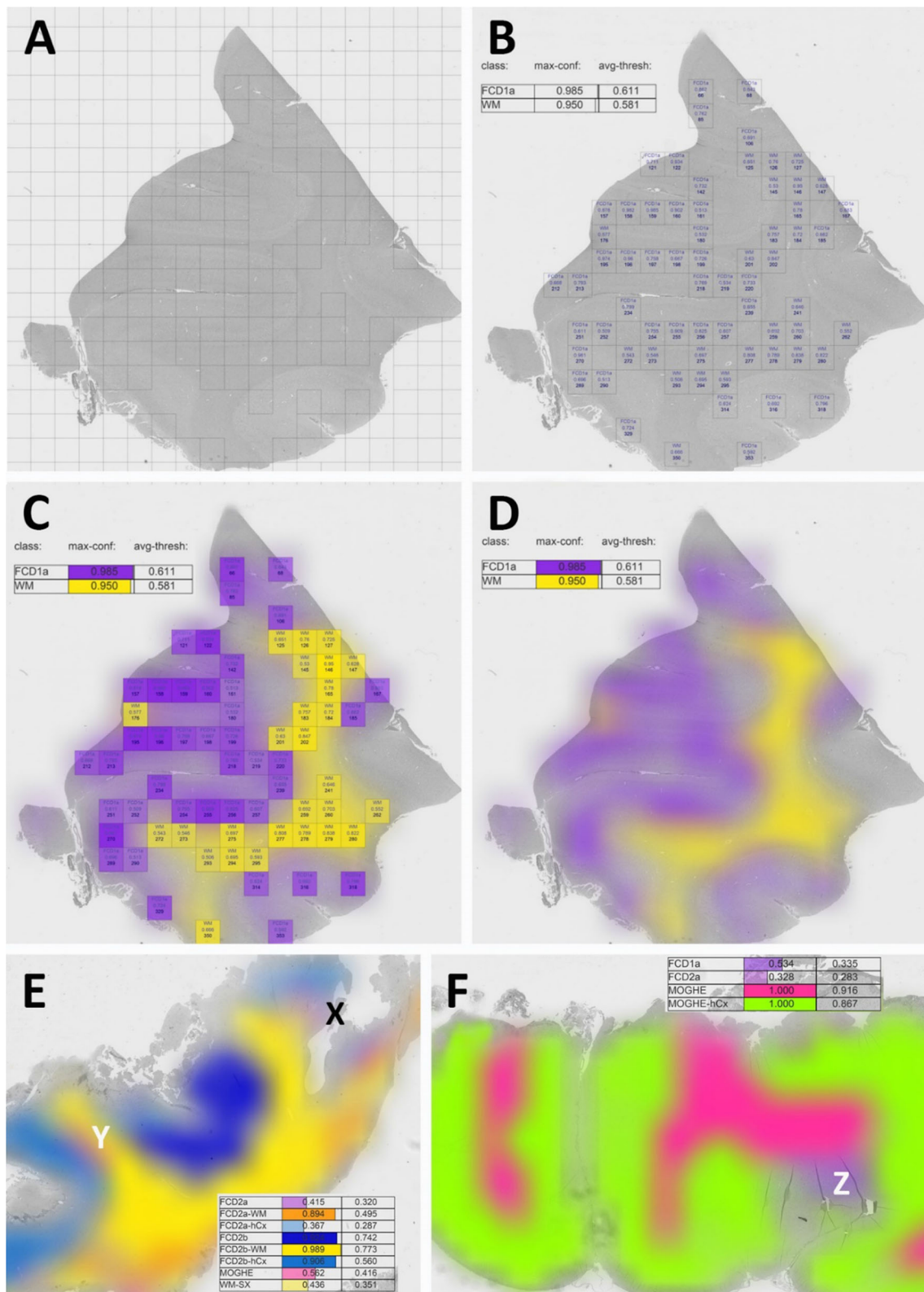


Fig. 2 Evaluation and visualization of a complete WSI slide. **a** A grid of tiles each measuring $1000 \times 1000 \mu\text{m}$, 1024×1024 pixel, respectively, was automatically calculated to cover the entire WSI image. **b** Prediction values for each tile were obtained from the trained CNN. The threshold to visualize the prediction value was defined as variable and set to 0.5 in this example. **c** All tiles reaching the defined threshold can be visualized by our color index map. Furthermore, the color intensity indicates the probability value. (The darker the color, the more probable is the prediction and vice versa.) Values were given on the upper left corner of the output file. **d** We used a blurring image filter for the final output image with all information shown in **a–c** hidden. Please note that the user can choose which of the 26 labels should be displayed, e.g., NTU is not shown in this example. **e** A validation case example of a WSI obtained from a patient with histopathologically confirmed FCD2b and genetically confirmed pathogenic MTOR mosaicism. Color coding of categories as follows: FCD2b in dark blue (max. confidence = 0.924); FCD 2b white matter in yellow (max. confidence = 0.969); FCD 2b neocortex in light blue (max. confidence = 0.367); FCD2a (X) and MOGHE (Y) were spotted across the WSI with low confidence (< 0.415 , 0.562) and were histopathologically not confirmed. **f** A validation case example of a WSI obtained from a patient with histopathologically confirmed MOGHE and SLC35A2 mutation. Red color coding for the prediction of MOGHE with a maximum confidence score of 1.000 and average confidence score of 0.916. The prediction score of FCD1a (Z) and FCD2a (not visible at upper right) was below 0.534 and 0.328, respectively. Other categories were not recognized herein

prediction threshold, and which were based on a pre-defined RGB color code table depending on the prediction probability as defined by the user (Fig. 2c). A thumbnail of the original WSI image was assembled from the individual png-formatted tiles, anticipating the large file size of each original WSI. The raster image was scaled up to the size of the generated thumbnail, and the two images were merged using an alpha channel and the pillow software package [10]. Our first visualization attempts confirmed the benefit of helper categories such as non-tissue background (NTU) and the subarachnoidal space (SUB-PM and SUB-SX) for training the CNN in order to improve the graphical output of the classifier. During each experiment, we applied the visualization program to 20 randomly selected cases of the validation cohort and microscopically reviewed regions of the slides identified by the CNN with an accuracy threshold above 0.5. This analog review confirmed the presence of the predicted lesion in all our cases, i.e., BA17, BA4, DG, mMCD, MOGHE, FCD1a, FCD2a and FCD2b. This system was further tested using WSI of genetically defined disease probes of MOGHE-*SLC35A2* altered, FCD2a-*DEPDC5* altered and FCD2b-*MTOR* altered. These histopathologically and genetically defined cases were not seen by the CNN before and were correctly assigned to their respective category (Fig. 2e, f).

A third approach addressed the potential disease differentiating anatomo-pathological signatures recognized by the CNN. We downloaded, therefore, the most accurately recognized tiles from the training and validation cohort of

the CNN with a predicted confidence ≥ 0.99 for microscopy review. It was evident from this analysis that the CNN has captured the specific disease tissue pattern for each category. Prominent examples are depicted in Fig. 3 showing the clusters of increased oligodendroglial cell densities in MOGHE, dysmorphic and balloon cells in FCD2b and excessive microcolumnar cortical architecture in FCD1a. An online accessible reviewing program was programmed and made accessible online to nine internationally renowned colleagues experienced in the classification of epilepsy-associated brain lesions to survey their agreement with our CNN's prediction. This study disclosed that 62.47% of the CNN's best tiles would have been classified as the respective brain disorder also by independent external referees, i.e., MOGHE, FCD1a, FCD2a, FCD2b (ranging from 10 to 96%, and kappa values ranging from poor (-0.80) to almost perfect (0.92), respectively; see Table 4).

Training results as well as the expert panel agreement trial suggested that our model was able to recognize complex tissue patterns within large WSI tiles with reasonable accuracy. To improve transparency of the model's decision-making process, we applied several discrimination software tools and algorithms for this interpretation using open access libraries, e.g., LIME [28], Captum [<https://doi.org/10.48550/arXiv.2009.07896>] with Guided-Grad-CAM [<https://doi.org/10.48550/arXiv.1610.02391>], integrated gradients [<https://doi.org/10.48550/arXiv.1703.01365>], an occlusion procedure [<https://doi.org/10.48550/arXiv.1311.2901>] as well as the torchcam library [github.com/frgfm/torch-cam] and the therein included Layer-CAM approach [<https://doi.org/10.1109/TIP.2021.3089943>] to visualize the common patterns of tissue texture as a heatmap trained in the model and which can help us guiding the microscopic examination of a surgical specimen [19] (Fig. 4).

We further investigated the class activation across convolution levels of our trained ResNet18 model applying the torchcam library and Layer-CAM approach included therein (Fig. 5). This analysis showed evidence for feature recognition taking place from layer 2 to layer 4 of the CNN. No more complex features were recognized in earlier or later convolutional layers of the model.

Finally, we assembled all features of our deep learning-based histopathology classifier for FCD into a web-based application to facilitate an automated interpretation of WSI slides of epilepsy surgery brain tissue (<https://fcd-classifier.eu.ngrok.io>). The application has several dashboards to access and review all of the information. It is important to note that the user can individually select the anatomical categories to be reviewed by the model, i.e., 1–25 (see glossary), and more importantly, that the user can verify the model's prediction ad hoc using a digital microscopy

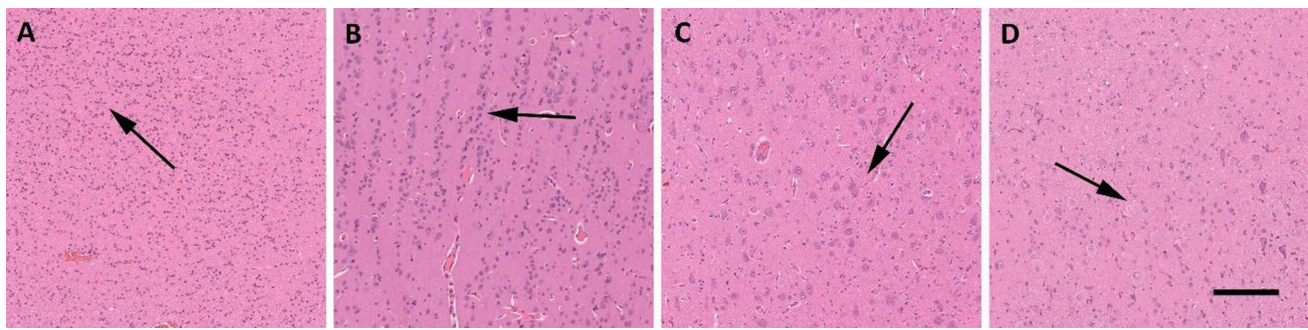


Fig. 3 Best tile prediction of the CNN with confidence scores above 0.99. **a** MOGHE with increased oligodendroglial cell density at the upper left (arrow). **b** FCD ILAE Type 1a with excessive microcolumns of the neocortex (arrow). **c** FCD ILAE Type 2a with

dysmorphic neurons (arrow). **d** FCD ILAE Type 2b with dysmorphic neurons and balloon cells (arrow). HE staining. Scale bar in D = 200 μ m applies also to A–C

Table 4 Independent expert review of 50 best tiles of each disease category predicted by the trained ResNet18 model

Reviewer	FCD1a	FCD2a	FCD2b	MOGHE	Total
A	70.00 (0.40)	92.00 (0.84)	96.00 (0.92)	84.00 (0.68)	85.50 (0.71)
B	18.00 (− 0.64)	80.00 (0.60)	60.00 (0.20)	50.00 (0.00)	52.00 (0.04)
C	10.00 (− 0.80)	96.00 (0.92)	52.00 (0.04)	64.00 (0.28)	55.50 (0.11)
D	46.00 (− 0.08)	60.00 (0.20)	38.00 (− 0.24)	38.00 (− 0.24)	45.50 (− 0.09)
E	24.00 (− 0.52)	42.00 (− 0.16)	32.00 (− 0.36)	38.00 (− 0.24)	34.00 (− 0.32)
F	42.00 (− 0.16)	42.00 (− 0.16)	34.00 (− 0.32)	40.00 (− 0.20)	39.50 (− 0.21)
G	68.00 (0.36)	78.00 (0.56)	84.00 (0.68)	22.00 (− 0.56)	63.00 (0.26)
H	32.00 (− 0.36)	76.00 (0.52)	72.00 (0.44)	66.00 (0.32)	61.50 (0.23)
J	92.00 (0.84)	94.00 (0.88)	68.00 (0.36)	70.00 (0.40)	81.00 (0.62)
Total	56.89 (0.23)	67.78 (0.36)	66.33 (0.33)	58.89 (0.18)	62.47 (0.25)

Nine independent experts (A–J) were invited to review 50 best tiles predicted by the model with a confidence score ≥ 0.99 . Please note that the agreement among experts varied substantially across all disease categories, i.e., from poor ($\kappa = -0.80$ /red) to almost perfect ($\kappa = 0.84$ /green) for FCD1a, from $\kappa = -0.16$ to $\kappa = 0.92$ for FCD2b, from $\kappa = -0.36$ to $\kappa = 0.92$ for FCD2a and from $\kappa = -0.56$ to 0.68 for MOGHE. Total—average across all values per category (columns) and across experts (rows). Each first number = percentage of agreed diagnosis from 50 best tiles. Number in brackets = calculated kappa value and assigned as following: POOR < 0.00; SLIGHT = 0.00–0.20; FAIR = 0.21–0.40; MODERATE = 0.41–0.60; SUBSTANTIAL = 0.61–0.80; ALMOST PERFECT = 0.81–1.00 [21]

tool. Figures 6, 7, 8 and 9 will help to guide through the application and get acquainted with its functionality. However, this version has a database of readily uploaded WSI images as the model can currently process only ndpi-image formats obtained from 4–5 μ m thin and HE-stained FFPE sections.

The possibility to review each tile's prediction score at the microscopic level is of utmost importance to validate the AI-based approach and to achieve a final and reliable histopathology diagnosis in routine practice.

4 Discussion

We developed a WSI-based deep learning algorithm to automatically detect and further classify FCD in epilepsy surgery brain specimen, including those entities newly

introduced in to the ILAE consensus FCD classification update of 2022, i.e., mMCD and MOGHE [22]. Comprehensive training and validation experiments revealed a success rate for the prediction of a total of 25 categories at 98.9%. The predicted categories can then be visualized at a high spatial resolution on the WSI by color coding and additional metrics to further specify these results. This prototype can be used online for diagnostic routine assessments and may become a template for a digital slide suite for neuropathology and epilepsy surgery in the near future. Notwithstanding, such an automated diagnostic device shall only support the diagnostic workup as a screening tool, e.g., when large surgically resected *en bloc* brain samples were submitted for neuropathology examination, and will not replace the microscopic inspection by an experienced and board certified neuro-/pathologist.

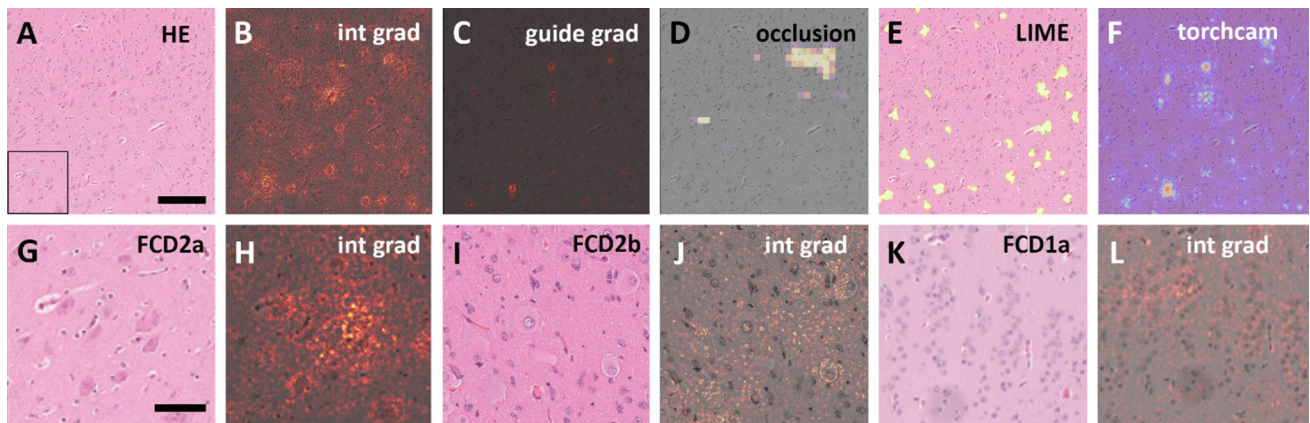


Fig. 4 Automated pattern recognition of tile content and its tissue texture. Different approaches for the interpretation of the classifier's prediction. From left to right: **a** original HE WSI tile of histopathologically confirmed FCD2a and recognized as FCD2a by our classifier as base; **b** Captum integrated gradients; **c** Captum with guided gradients, B and C recognized dysmorphic neurons at different sensitivity levels; **d** an occlusion algorithm visualized no specific cellular pattern and was not further investigated; **e** interpretation of

the discriminating image area for the classifiers' prediction based on the python library LIME; and **f** torchcam library (see also Fig. 5). **g, h** Higher magnification of same HE WSI shown above with integrated gradient interpretation recognizing dysmorphic neurons of FCD2a; **i–j** integrated gradient interpretation recognizing balloon cells of FCD2b; **k–l** integrated gradient interpretation recognizing microcolumns of FCD1a. Color coding of the interpretation maps: dark = low with min. of 0.0, light = high with max of 1.0

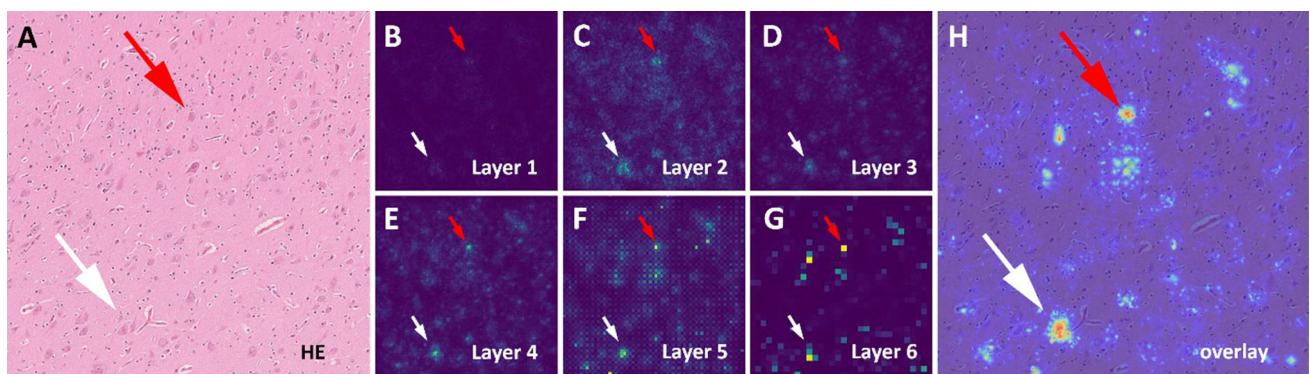


Fig. 5 Layered ResNet18 activation of complex tissue patterns. **a** Input image at 1024×1024 pixel. **b** Activation of convolutional layer 1. No features were detected by the network (512×512 pixel); **c** Pattern recognition was strongest at layer 2 as indicated by a color scale from dark (0.0) to light (1.0), size of activation map 256×256 pixel; and **d, e** No detection of additional complex features in the following network layers 3 and 4. **f, g** Reinforcement of activation in

already recognized image areas through interpolation of the decreasing activation maps in the course of the model, i.e., layers 5 and 6 (64×64 pixel and 32×32 pixel, respectively). **h** Superimposed image of all layer activations highlighted predominantly areas of dysmorphic neurons as confirmed by the input image in A. White and red arrows in A–H repeatedly point to same two regions with dysmorphic neurons as recognized by the model

To the best of our knowledge, our deep learning-based histopathology classifier of FCD is first to process large image tiles of $1000 \times 1000 \mu\text{m}$. Digital tiles of this size represent a rich brain tissue texture with complex cellular and extracellular content, e.g., the various neuronal profiles of different diameter and shape, orientation and architectural organization in layers. The same holds true for the various glial cell types, endothelial cells, blood vessel diameter and the large extracellular space comprising almost 20% of the adult brain tissue [25, 27]. In addition, perisurgical and laboratory artifacts, e.g., bleeding, wrinkles and scratches, were often included in such large tiles and also challenge this approach. During the experimental

training phases, the question arose whether a 1 mm^2 tile size is suited for this classifier challenge as all previous models have used smaller images focusing only on the disease-specific cellular pattern, e.g., dysmorphic neurons w/o calcification in cortical tubers vs. FCD2b [8, 19] as well as addressing specific histopathology domains, such as carcinoma islets [8], pituitary glands or immunohistochemical staining patterns [24]. We computed smaller tiles, therefore, during the course of this project, e.g., of 224×224 and 512×512 pixel. Accuracy scores were lower in all of these conditions (data not shown), and we concluded from these results the algorithm's difficulty to

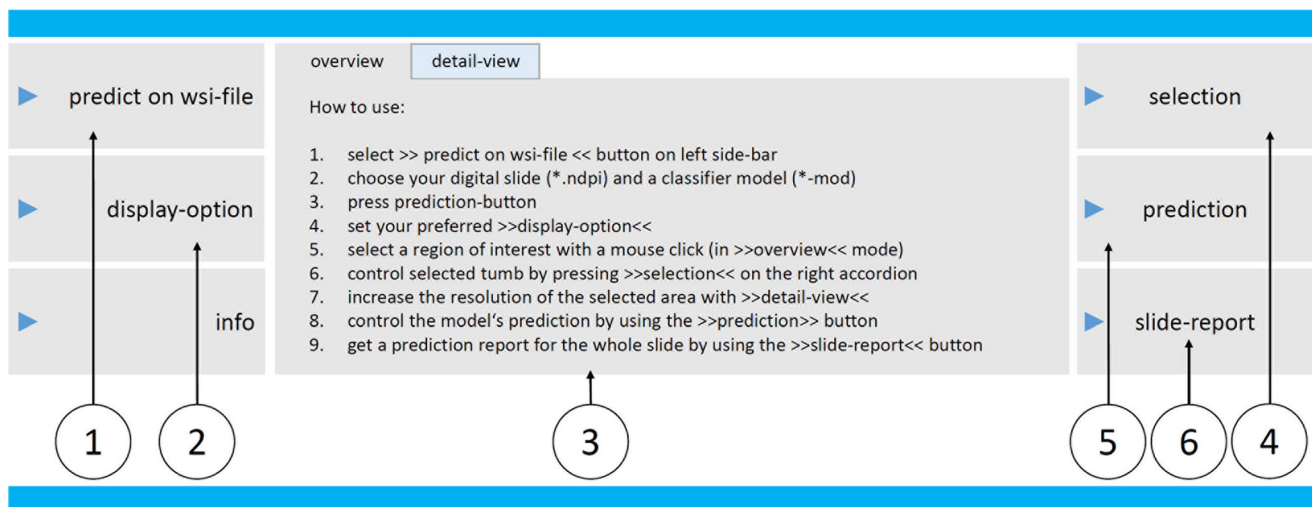


Fig. 6 Landing page of the online FCD classifier application. **1** First, go to the accordion-tab on the left to select your WSI of interest (see also Fig. 7). A set of WSI slides is already stored in its database. Future versions will enable the upload of external WSI. Technical issues to allow different WSI formats and size need to be adjusted, however. This tab also allows to choose among trained classifier models. Currently, the herein described basic ResNet18 and One-vs-All FCD2b models can be selected (supplement Table 2), but future applications may envisage different topic-specific models addressing

the larger spectrum of lesions associated with epilepsy surgery [6]. **2** Adjust the visual representation of the classifier's prediction (see Fig. 8 for more detail). **3** Output area: at this entry level it provides a short instruction manual. At other stages it reveals the status of the system during/after the calculation process, or will allow zooming, panning and saving of the output image. **4** Access to live microscopy review of a chosen area of interest (HE-stained WSI, Fig. 9). **5** To review the model's prediction scores of all chosen categories. **6** Receive a report

differentiate FCD while admixed with normal brain tissue of the neocortex and white matter.

We reviewed the microcopy signature of best predicted tiles chosen by the CNN (> 0.99) for each category and recognized the prototypical tissue texture (Fig. 3). In addition, we applied several discrimination software tools to highlight the CNN's activation using open access libraries (Fig. 4). Results confirmed the recognition of prototypical tissue texture including cellular disease hallmarks, e.g., dysmorphic neurons in FCD2a and FCD2b and excessive microcolumns in FCD 1a (Figs. 4, 5). This is an interesting observation given the similarity of many histopathology features and the variability of normal anatomical landmarks. As an example, the microcolumnar architecture defining FCD1a was also reported to be partially presented in normal brain, in particular in the occipital region close or adjacent to BA17 [16]. Giant cells of Betz characterize Brodmann area 4 in the frontal lobe and may be difficult to differentiate from dysmorphic neurons in FCD2 of the frontal lobe (supplemental Figure x1A). Our trained application never reported this issue, however. White matter was always differentiated from neocortex based on HE staining alone, and the edges of the tissue specimen were always correctly recognized by the model, i.e., the subarachnoidal space. The assignment of different white matter categories still poses the biggest challenge to the system (Table 3), whether it belongs to normal brain obtained from patients without any epileptic

disorders or to mMCD characterized by an excess of heterotopic neurons in the white matter or to MOGHE with increased oligodendroglial cell densities. Results predicted by the system should, therefore, always be confirmed by microscopic inspection of an experienced and specialized neuro-/pathologist and preferably also by specific immunohistochemical markers as recommended by the ILAE classification schemes [2, 7].

We took this approach to another level and invited nine histopathologists around the world and experienced in epilepsy surgery to review the 50 best tiles selected by the CNN (Table 4). This trial was not designed to challenge the CNN's selection but rather understand its prediction rules. The results were stunning and very much comparable to the variable inter-rater agreement of previous histopathology trials in epilepsy surgery when presenting analog glass slides or digital WSI series to be iteratively reviewed by expert panels [4, 9, 11]. Such agreement scores reached 62.5% in one study [9], with best values for FCD Type 2 and lowest for FCD Type 1, respectively. Similarly, a kappa value of 0.64 was reported for the FCD classification scheme of 2011 with best consensus for FCD Type 2 categories [11]. A more recent approach achieved a kappa score of 0.65 only after four rounds of microscopic review and when molecular testing was disclosed [4]. These scores were pretty similar to those obtained in our current study, however. Overall, our inter-rater agreement varied between 10 and 96% across the expert panel with kappa values

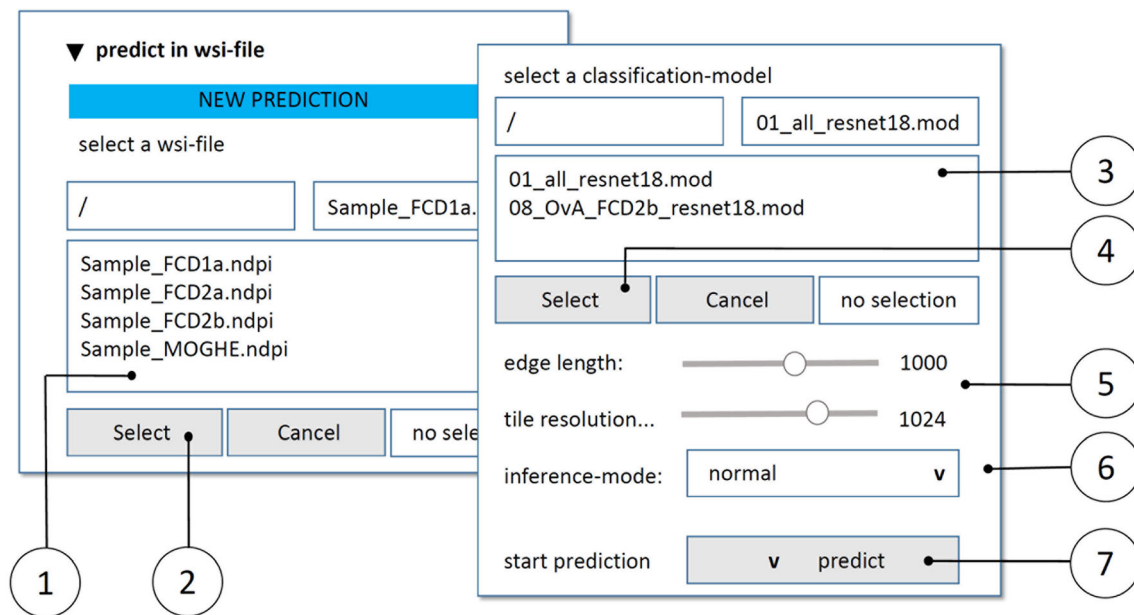


Fig. 7 How to pick your slide and DL model. Please click on the “predict on wsi file” accordion in the upper left to start the application. Two file selection dialogs will be presented in order to choose a WSI file (shown on left) and a trained prediction model (shown on right). The application already provides WSI files obtained from epilepsy surgery as well as two trained models: 01_all_resnet18.mod representing our basic model (Table 3); the second model represents the One-vs-All model for FCD2b (supplement Table 2). File upload is not enabled in this version. In addition, the selection list is secured by an informatics sandbox. Take the following steps to complete this task. **1** Select your slide of interest. **2** Push “select” to confirm your selection. **3** Select your DL model. **4**

Push “select” to confirm your selection. **5** You can use sliders to manually adapt the parameters of the image. Please note that dimension and resolution must match that of the trained model, i.e., 1024×1024 pixels and image edge length of 1000 mm. These are set as default and should not be changed. **6** Choose your inference mode. In “normal” mode, each extracted tile is scored simple. In the “test-time augmentation” (tta) mode, each tile is evaluated four times to further augment the image output. This procedure takes significantly longer in time. **7** Push the “predict” button to start the algorithm. Calculation progress can be followed on the output area (Fig. 6)

ranging from poor (-0.80) to almost perfect (0.96). The category with best agreement was that of FCD2a, for which three raters reached an almost perfect kappa score of > 0.81 . In contrast, the FCD1a category achieved agreement with the CNN’s selection in 10% to 92% of tiles and kappa values ranging from -0.80 (poor) to 0.84 (almost perfect). While these numbers tell us that some raters could recognize the features selected by the CNN, the subjective nature of human microscopy studies in FCD remains an ever-challenging issue. This assumption also supports our foremost study aim to develop an AI-based histopathology classifier to support difficult-to-diagnose FCD lesions at the microscopy level.

Another application of this model could be the recognition of double pathologies which is currently not included in the ILAE classification of FCD. Although we have not systematically performed such studies with this new classifier, the probability that architectural dysplasia of FCD ILAE Type 1 can occur in the vicinity of FCD Type 2 lesions caught our attention in several cases (see Fig. 2f) and shall be further examined in the near future. Another challenge remains, however, that the model falsely assigns

small regions or individual tiles to a different disease category when the majority of tiles were predictive otherwise. An example is also shown in Fig. 2e, f with genetically confirmed FCD2b or MOGHE and small islets of other FCD entities identified herein by our model. It is important to review the prediction score, which was below 0.5 in both cases confirming a low probability of such a double pathology. Visual inspection of the microscopic slide did also not confirm the presence of additional pathologies. We recommend to use this classifier, therefore, as a screening tool in which areas with a high probability rate should be microscopically reviewed to achieve a final diagnosis.

Despite the great success in the application of CNN in the arena of digital pathology, we have to also consider its limitation. Poor traceability of the decision-making process is due to the complexity of CNN architectures. We addressed this issue (see Fig. 4) but cannot further resolve the nature of its decision-making, yet. Another limitation in the context of digital pathology reflects on the robustness of generalization. We cannot exclude overfitting from use of a single scanner model and protocols established in our histopathology laboratory, including formalin fixation,

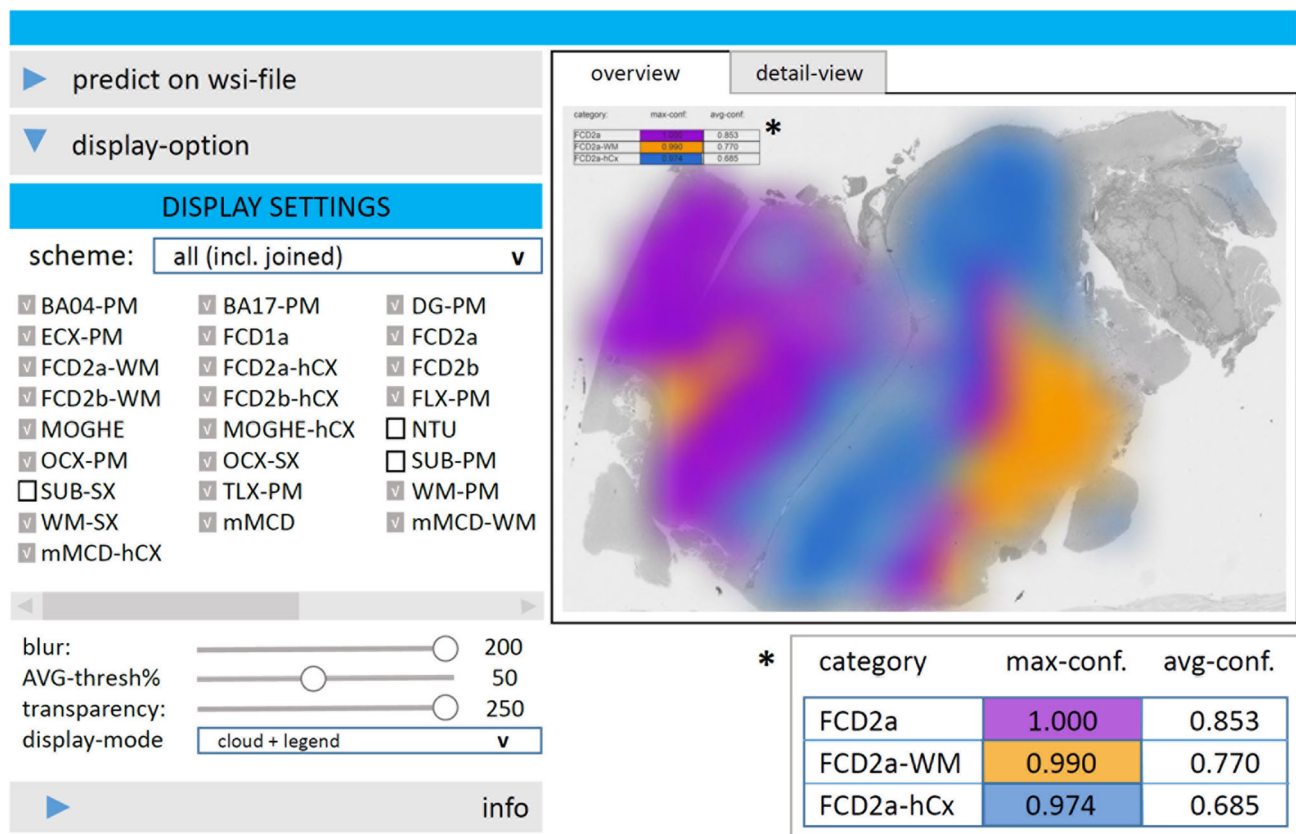


Fig. 8 Use various display options to visualize the results. The box “overview” of the middle panel will reveal the computed prediction map for the WSI. The accordion on the left allows to modify the display options by selecting either all anatomic categories available for the model or a preselection thereof. Check boxes on left to make a choice. Please note that only checked as well as recognized categories will be shown. For your convenience, maximum and average confidence scores for those categories detected are readily visible on the upper left of the image. The sliders on the lower left allow to adjust the graphic presentation with more or less “blurring” to visualize all tiles. “Transparency” is a measure of the detection probability. The AVG-threshold slider can be applied to define or modify the prediction threshold. Only tiles above this threshold will

then be shown in the image. The last tab on the lower left allows the user to choose the display mode. The application offers different illustrations of color-coded tiles, e.g., as clouds, rectangles or circles, each with/out recognition scores as attached legend (see supplement Figure 2). When color-coded circles are chosen, the diameter of each circle indicates the level of the detection probability. The output image will be recalculated and displayed in the middle panel following each change of display options. Please accept any time delay of the output results after each interactive adaptation of the user due to data transfer and recalculation (> 5–10 s). Results will best displayed at 16:9 scaled monitors. Asterisk in overview: the output table is magnified on lower right for better readability

paraffin embedding, tissue cutting and use of staining dyes (Schmitt et al. 2021; <https://doi.org/10.2196/23436>). These issues need to be addressed in future studies when testing the model with datasets from different laboratories. Previous experiments showed also that CNN networks are vulnerable to adversarial attacks, when smallest permutations at the pixel level of an input image can cause a change in the predicted value of a classification model (Su et al. 2017; <https://doi.org/10.48550/arXiv.1710.08864>). Other approaches have used artificially produced images that are completely unrecognizable to humans, but can force CNN’s to predict outcomes with high confidence, (Nguyen et al. 2014; <https://doi.org/10.48550/arXiv.1412.1897>, Kumano et al. 2020; <https://doi.org/10.48550/arXiv.2012.03843>).

This implies that a trained classification model should only be used with image inputs for which it was explicitly trained to avoid misclassification. In addition, a basic knowledge of the structure and functionality of the classification network is necessary to ensure its correct handling and interpretation of results. This notion demands data preselection and coarse pre-classification by a pathology expert in the context of digital neuropathology of epilepsy surgery. As mentioned above, we unanimously mandate to use such tools in the future as diagnostic aid to reduce work burden but not to replace the diagnostic decision maker.

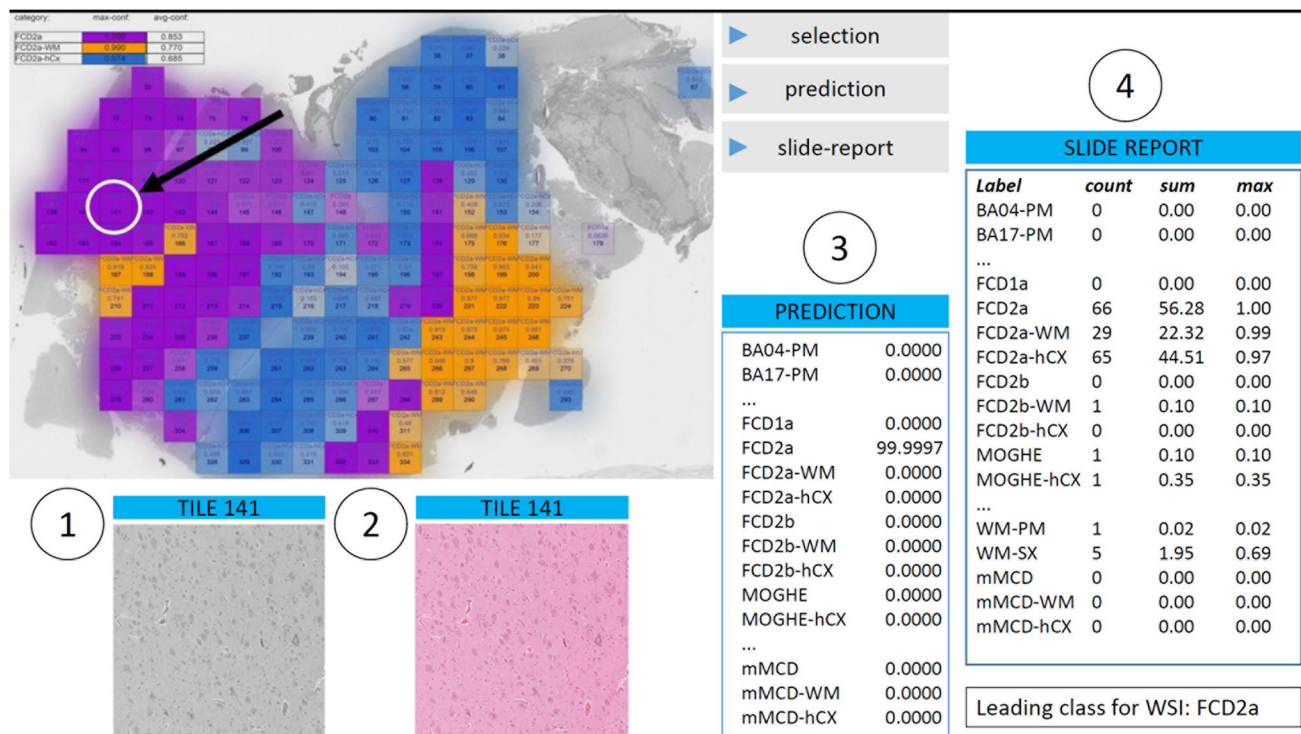


Fig. 9 Review the results. This figure describes the functionality of the three tabs on the upper right of the application (see Fig. 6), i.e., selection, predictions and slide report in rectangle mode with legend display. You can review the original histopathology image (WSI slide) of each tile to confirm or challenge the model's prediction. Click on any region of interest on the output image, e.g., tile 141 (indicated by white circle on output image). **1** A gray-colored thumbnail with a resolution of 1024×1024 pixels will be displayed for this tile. **2** If the <<detail-view>> button is pressed, this tile is retrieved from the WSI file and the HE staining accessible at high resolution for the user's digital microscopy review. **3** The prediction

scores for the selected tile were disclosed for all categories (... to secure space on this image we removed few categories without any further information). **4** A slide report can be issued for the studied WSI: 1st column: selected categories (label); 2nd column: number of tiles assigned to this class (count); 3rd column: sum of predicted scores of all tiles assigned to this category (sum); 4th column: highest prediction score (0–1) calculated for any tile of this category (max); and 5th column (not shown): average value of all predictions across all tiles of this category (avg). The footer displays the leading class (category) for this WSI. In this example, the model predicts FCD2a with an average prediction score of 85%

4.1 A free access, online application of the FCD classifier

We developed a web-based open access application to use our FCD classifier on a daily diagnostic routine basis. This application may also become helpful in view of the first update of the FCD consensus classification scheme of the ILAE [22]. This update includes for the first time the histopathologically difficult-to-diagnose disease entities of mMCD and MOGHE, and our model can help to train colleagues not yet familiar with these diagnosis online. The application prototype is accessible via <https://fcd-classifier.eu.ngrok.io>. A user will have to register to the platform to get free access (please send an email with your registration request to the senior authors). Currently, the application has a database of securely stored WSI for testing purpose as our analysis pipeline is able to compute only WSI of the Hamamatsu scanning.ndpi/.ndpa formats, however. Further advancements to add conventional WSI formats are in progress but largely pending on the availability of a

common DICOM format offered from all vendors [18]. It will also be mandatory that the FFPE tissue section has a cutting thickness of 4–5 μm and the HE staining protocol was performed according to ILAE recommendations [2]. The large WSI file size is another hurdle for any transfer and upload pipeline and should be limited, therefore, to < 3 Gb.

In conclusion, we developed an AI-based histopathology classifier to support routine diagnosis of FCD samples obtained from epilepsy surgery. The ResNet18-trained classifier currently recognizes 25 anatomo-pathological categories of five major disease entities compared to normal human neocortex. Major achievements include the use of large 1 mm^2 WSI tiles allowing to recognize the disease phenotype admixed with an otherwise normal appearing human neocortex or white matter. Extraction of the activation patterns showed concordance with disease defining cellular phenotypes which was also confirmed by our invited expert panel. However, the low concordance among histopathology experts is an ongoing challenge in the

neuropathological examination of epilepsy surgery brain tissue specimens and may be resolved by introducing our AI-based algorithm available online and open access to the worldwide community of neuro-/pathologists.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00521-023-08364-9>.

Acknowledgements The present work was performed in fulfillment of the requirements of the Friedrich-Alexander Universität Erlangen-Nürnberg (FAU) for obtaining the degree “Dr. rer. biol. hum.” of Jörg Vorndran. We kindly thank our colleagues from Bielefeld-Bethel, Munich and Vogtareuth for sharing their tissue samples and clinical information (see supplement Table x1). Histopathologists from Mexico City, Paris, Belgrade, Milano, Akita, Utrecht, Kuopio and Campinas participated in our agreement trial. The Erlangen histopathology team was instrumental to retrieve the samples, re-cut and stain missing/damaged slides, as well as digitizing and annotating all specimens.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest None of the authors have any disclosure.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Blümcke I, Aronica E, Becker A, Capper D, Coras R, Honavar M, Jacques TS, Kobow K, Miyata H, Mühlebner A et al (2016) Low-grade epilepsy-associated neuroepithelial tumours—the 2016 WHO classification. *Nat Rev Neurol* 12:732–740
- Blumcke I, Aronica E, Miyata H, Sarnat HB, Thom M, Roessler K, Rydenhag B, Jehi L, Krsek P, Wiebe S et al (2016) International recommendation for a comprehensive neuropathologic workup of epilepsy surgery brain tissue: a consensus Task Force report from the ILAE Commission on Diagnostic Methods. *Epilepsia* 57:348–358
- Blumcke I, Budday S, Poduri A, Lal D, Kobow K, Baulac S (2021) Neocortical development and epilepsy: insights from focal cortical dysplasia and brain tumours. *Lancet Neurol* 20:943–955. [https://doi.org/10.1016/S1474-4422\(21\)00265-9](https://doi.org/10.1016/S1474-4422(21)00265-9)
- Blumcke I, Coras R, Busch RM, Morita-Sherman M, Lal D, Prayson R, Cendes F, Lopes-Cendes I, Rogerio F, Almeida VS et al (2021) Toward a better definition of focal cortical dysplasia: an iterative histopathological and genetic agreement trial. *Epilepsia* 62:1416–1428. <https://doi.org/10.1111/epi.16899>
- Blumcke I, Coras R, Wefers AK, Capper D, Aronica E, Becker A, Honavar M, Stone TJ, Jacques TS, Miyata H et al (2019) Challenges in the histopathological classification of ganglioglioma and DNT: microscopic agreement studies and a preliminary genotype-phenotype analysis. *Neuropathol Appl Neurobiol* 45:95–107. <https://doi.org/10.1111/nan.12522>
- Blümcke I, Spreafico R, Haaker G, Coras R, Kobow K, Bien CG, Pfafflin M, Elger C, Widman G, Schramm J et al (2017) Histopathological findings in brain tissue obtained during epilepsy surgery. *N Engl J Med* 377:1648–1656. <https://doi.org/10.1056/NEJMoa1703784>
- Blümcke I, Thom M, Aronica E, Armstrong DD, Vinters HV, Palmini A, Jacques TS, Avanzini G, Barkovich AJ, Battaglia G et al (2011) The clinico-pathological spectrum of Focal Cortical Dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia* 52:158–174
- Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ (2019) Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 25:1301–1309. <https://doi.org/10.1038/s41591-019-0508-1>
- Chamberlain WA, Cohen ML, Gyure KA, Kleinschmidt-DeMasters BK, Perry A, Powell SZ, Qian J, Staugaitis SM, Prayson RA (2009) Interobserver and intraobserver reproducibility in focal cortical dysplasia (malformations of cortical development). *Epilepsia* 50:2593–2598
- Clark A (2015) Pillow (pil fork) documentation. Readthedocs <https://www.BuildmediaReadthedocs.Org/Media/Pdf/Pillow/Latest/Pillow.pdf>
- Coras R, de Boer OJ, Armstrong D, Becker A, Jacques TS, Miyata H, Thom M, Vinters HV, Spreafico R, Oz B et al (2012) Good interobserver and intraobserver agreement in the evaluation of the new ILAE classification of focal cortical dysplasias. *Epilepsia* 53:1341–1348
- de Haan K, Zhang Y, Zuckerman JE, Liu T, Sisk AE, Diaz MFP, Jen KY, Nobori A, Liou S, Zhang S et al (2021) Deep learning-based transformation of H&E stained tissues into special stains. *Nat Commun* 12:4884. <https://doi.org/10.1038/s41467-021-25221-2>
- Engel J (1987) Outcome with respect to epileptic seizures. In: Engel JJ (ed) *Surgical treatment of the Epilepsies*. Raven, City, pp 553–571
- Ghaffari Laleh N, Muti HS, Loeffler CML, Echle A, Saldanha OL, Mahmood F, Lu MY, Trautwein C, Langer R, Dislich B et al (2022) Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology. *Med Image Anal* 79:102474. <https://doi.org/10.1016/j.media.2022.102474>
- Gill RS, Lee HM, Caldaïrou B, Hong SJ, Barba C, Deleo F, D’Incerti L, Mendes Coelho VC, Lenge M, Semmelroch M et al (2021) Multicenter validation of a deep learning detection algorithm for focal cortical dysplasia. *Neurology* 97:e1571–e1582
- Hildebrandt M, Pieper T, Winkler P, Kolodziejczyk D, Holthausen H, Blumcke I (2005) Neuropathological spectrum of cortical dysplasia in children with severe focal epilepsies. *Acta Neuropathol* 110:1–11
- Holthausen H, Coras R, Tang Y, Bai L, Wang I, Pieper T, Kudernatsch M, Hartlieb T, Staudt M, Winkler P et al (2022) Multilobar unilateral hypoplasia with emphasis on the posterior

- quadrant and severe epilepsy in children with FCD ILAE Type 1A. *Epilepsia* 63:42–60. <https://doi.org/10.1111/epi.17114>
18. Kolble K, Cross JH, Becker A, Blumcke I (2018) A web-based diagnostic reference centre for the European Reference Network “EpiCare”: recommendations of the eNeuropathology working group. *Epileptic Disord* 20:339–345. <https://doi.org/10.1684/epd.2018.1002>
 19. Kubach J, Muehlechner-Fahrngruber A, Soylemezoglu F, Miyata H, Niehusmann P, Honavar M, Rogerio F, Kim SH, Aronica E, Garbelli R et al (2020) Same same but different: a Web-based deep learning application revealed classifying features for the histopathologic distinction of cortical malformations. *Epilepsia* 61:421–432. <https://doi.org/10.1111/epi.16447>
 20. Lamberink HJ, Otte WM, Blümcke I, Braun KPJ (2020) Seizure outcome and use of antiepileptic drugs after epilepsy surgery according to histopathological diagnosis: a retrospective multi-centre cohort study. *Lancet Neurol* 19:748–757. [https://doi.org/10.1016/s1474-4422\(20\)30220-9](https://doi.org/10.1016/s1474-4422(20)30220-9)
 21. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
 22. Najm I, Lal D, Vanegas MA, Cendes F, Lopes-Cendes I, Palmini A, Paglioli E, Sarnat H, Walsh CA, Wiebe S et al (2022) The ILAE consensus classification of Focal Cortical Dysplasia (FCD): an update proposed by an ad hoc Task Force of the ILAE Diagnostic Methods commission. *Epilepsia* 63:1899–1919
 23. Najm IM, Sarnat HB, Blümcke I (2018) The international consensus classification of Focal Cortical Dysplasia—a critical update 2018. *Neuropathol Appl Neurobiol* 44:18–31. <https://doi.org/10.1111/nan.12462>
 24. Neuner C, Coras R, Blümcke I, Popp A, Schlaffner SM, Wirries A, Buchfelder M, Jabari S (2022) A whole-slide image managing library based on Fastai for deep learning in the context of histopathology: two use-cases explained. *Appl Sci* 12:13
 25. Ohashi T, Edamatsu M, Bekku Y, Carulli D (2015) The hyaluronan and proteoglycan link proteins: organizers of the brain extracellular matrix and key molecules for neuronal function and plasticity. *Exp Neurol* 274:134–144. <https://doi.org/10.1016/j.expneurol.2015.09.010>
 26. Pedregosa F, Varoquaux G, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J et al (2011) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
 27. Rauch U (2004) Extracellular matrix components associated with remodeling processes in brain. *Cell Mol Life Sci* 61:2031–2045. <https://doi.org/10.1007/s00018-004-4043-x>
 28. Ribeiro MT, Singh S, Guestrin C (2016) “Why should I trust you?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. association for computing machinery, City, pp 1135–1144
 29. Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y, Gunaydin H, Zuckerman JE, Chong T, Sisk AE et al (2019) Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng* 3:466–477. <https://doi.org/10.1038/s41551-019-0362-y>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Jörg Vorndran^{1,20} · Christoph Neuner^{1,20} · Roland Coras^{1,20} · Lucas Hoffmann^{1,20} · Simon Geffers¹ · Jonas Honke¹ · Jochen Herms² · Sigrun Roeber² · Hajo Hamer^{3,20} · Sebastian Brandner^{4,20} · Till Hartlieb^{5,6,20} · Tom Pieper⁶ · Manfred Kudernatsch^{6,7} · Christian G. Bien⁸  · Thilo Kalbhenn⁹ · Matthias Simon⁹ · Homa Adle-Biasette^{10,11} · Jesús Cienfuegos¹² · Roberta Di Giacomo^{13,20} · Rita Garbelli^{13,20} · Hajime Miyata¹⁴  · Angelika Muehlechner^{15,20} · Savo Raicevic¹⁶ · Tuomas Rauramaa^{17,20}  · Fabio Rogerio^{18,19} · Ingmar Blümcke^{1,20}  · Samir Jabari^{1,20}

✉ Ingmar Blümcke
bluemcke@uk-erlangen.de

¹ Department of Neuropathology, Universitätsklinikum Erlangen, Erlangen, Germany

² Zentrum für Neuropathologie, Ludwig-Maximilian Universität München, Munich, Germany

³ Epilepsy Center, Universitätsklinikum Erlangen, FAU Erlangen-Nürnberg, Erlangen, Germany

⁴ Department of Neurosurgery, Universitätsklinikum Erlangen, Erlangen, Germany

⁵ Center for Pediatric Neurology, Neurorehabilitation and Epileptology, Schoen Klinik Vogtareuth, Vogtareuth, Germany

⁶ Research Institute Rehabilitation, Transition, Palliation, Paracelsus Medical University Salzburg, Salzburg, Austria

⁷ Center for Neurosurgery, Epilepsy Surgery, Spine Surgery and Scoliosis, Schoen Klinik Vogtareuth, Vogtareuth, Germany

⁸ Department of Epileptology (Krankenhaus Mara), Medical School, Universitätsklinikum Ostwestfalen-Lippe, Bielefeld, Germany

⁹ Department of Neurosurgery (Evangelisches Klinikum Bethel), Medical School, Universitätsklinikum Ostwestfalen-Lippe, Bielefeld, Germany

¹⁰ Université de Paris Cité, NeuroDiderot, Inserm, Paris, France

¹¹ Service d'Anatomie Pathologique, Hôpital Lariboisière, AP-HP, Paris, France

¹² International Center for Epilepsy Surgery, Hospital HMG, México City, Mexico

¹³ Epilepsy Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milano, Italy

- ¹⁴ Department of Neuropathology, Research Institute for Brain and Blood Vessels, Akita Cerebrospinal and Cardiovascular Center, Akita, Japan
- ¹⁵ UMC Utrecht Brain Center, Department of (Neuro)pathology, University Medical Center Utrecht and Utrecht University, Utrecht, Netherlands
- ¹⁶ Laboratory for Neuropathology, Department of Pathology, Clinical Centre of Serbia, Belgrade, Serbia
- ¹⁷ Department of Pathology, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland
- ¹⁸ Department of Pathology, University of Campinas, Sao Paulo, Brazil
- ¹⁹ Brazilian Institute of Neuroscience and Neurotechnology, Sao Paulo, Brazil
- ²⁰ Present Address: Partner of the European Reference Network (ERN), EpiCare, Lyon, France