

UC Davis

UC Davis Previously Published Works

Title

A machine learning approach to automate microinfarct and microhemorrhage screening in hematoxylin and eosin-stained human brain tissues

Permalink

<https://escholarship.org/uc/item/1cf9z8dg>

Journal

Journal of Neuropathology & Experimental Neurology, 84(2)

ISSN

0022-3069

Authors

Oliveira, Luca Cerny

Chauhan, Joohi

Chaudhari, Ajinkya

et al.

Publication Date




2024-12-26

DOI

10.1093/jnen/nlae120

Peer reviewed

A machine learning approach to automate microinfarct and microhemorrhage screening in hematoxylin and eosin-stained human brain tissues

Luca Cerny Oliveira , MS¹, Joochi Chauhan, PhD^{1,2}, Ajinkya Chaudhari, BS¹, Sen-ching S. Cheung, PhD³, Viharkumar Patel, MD², Amparo C. Villablanca, MD⁴, Lee-Way Jin, MD, PhD², Charles DeCarli, MD², Chen-Nee Chuah , PhD¹, Brittany N. Dugger , PhD^{*,2}

¹Department of Electrical and Computer Engineering, University of California Davis, Davis, CA, United States

²Department of Pathology and Laboratory Medicine, University of California Davis, Sacramento, CA, United States

³Department of Electrical and Computer Engineering, University of Kentucky, Lexington, KY, United States

⁴Department of Internal Medicine, University of California Davis, Davis, CA, United States

*Send correspondence to: Brittany N. Dugger, PhD, Department of Pathology and Laboratory Medicine, UC Davis School of Medicine, 4645 2nd Ave., 3400A Research Building III, Sacramento, CA 95817, United States; E-mail: bndugger@ucdavis.edu

Author Contributions: L. Cerny Oliveira and J. Chauhan shared first authorship.

ABSTRACT

Microinfarcts and microhemorrhages are characteristic lesions of cerebrovascular disease. Although multiple studies have been published, there is no one universal standard criteria for the neuropathological assessment of cerebrovascular disease. In this study, we propose a novel application of machine learning in the automated screening of microinfarcts and microhemorrhages. Utilizing whole slide images (WSIs) from postmortem human brain samples, we adapted a patch-based pipeline with convolutional neural networks. Our cohort consisted of 22 cases from the University of California Davis Alzheimer's Disease Research Center brain bank with hematoxylin and eosin-stained formalin-fixed, paraffin-embedded sections across 3 anatomical areas: frontal, parietal, and occipital lobes (40 WSIs with microinfarcts and/or microhemorrhages, 26 without). We propose a multiple field-of-view prediction step to mitigate false positives. We report screening performance (ie, the ability to distinguish microinfarct/microhemorrhage-positive from microinfarct/microhemorrhage-negative WSIs), and detection performance (ie, the ability to localize the affected regions within a WSI). Our proposed approach improved detection precision and screening accuracy by reducing false positives thereby achieving 100% screening accuracy. Although this sample size is small, this pipeline provides a proof-of-concept for high efficacy in screening for characteristic brain changes of cerebrovascular disease to aid in screening of microinfarcts/microhemorrhages at the WSI level.

KEYWORDS: deep learning; digital pathology; histology; infarcts; vascular dementia

INTRODUCTION

Microinfarcts and microhemorrhages within the brains of persons with cerebrovascular disease increase the risk of vascular dementia and are only discernible upon microscopic evaluation.¹ Microinfarcts/microhemorrhages can result from small vessel disease, arteriolosclerosis, and/or the rupture or obstruction of cerebral vessels.² There are various histological features of these microscopic lesions based on temporal aspects; they can be very heterogeneous.³ An overview of different types of these lesions is shown in Figure 1. Infarcts can vary by age (chronic, subacute, and acute),⁴ size (lacunar, micro, and macro/large),⁵ and may include hemorrhagic components. The is considerable heterogeneity which can be based on varying anatomic appearances, there have been

reports of variable interrater agreement.^{3,6} This complexity is compounded by the fact there are many proposed standardized criteria for neuropathological assessment of cerebrovascular diseases,^{6–11} with no single universal criteria.^{1,6–11}

Digital pathology, empowered by computational tools for medical image analysis, has revolutionized the characterization of diseases.¹² Machine learning (ML) enhances this capability by offering automated and scalable methods for examining pathologies in whole slide images (WSIs), the digitized version of microscope slides.^{13–15} Recent research demonstrates ML models have attained competitive results in the automated analysis of many pathologies in human post-mortem brains.^{16–21} Additionally, there have been successful in-vivo studies using ML to classify or segment vascular pathologies, such as infarcts

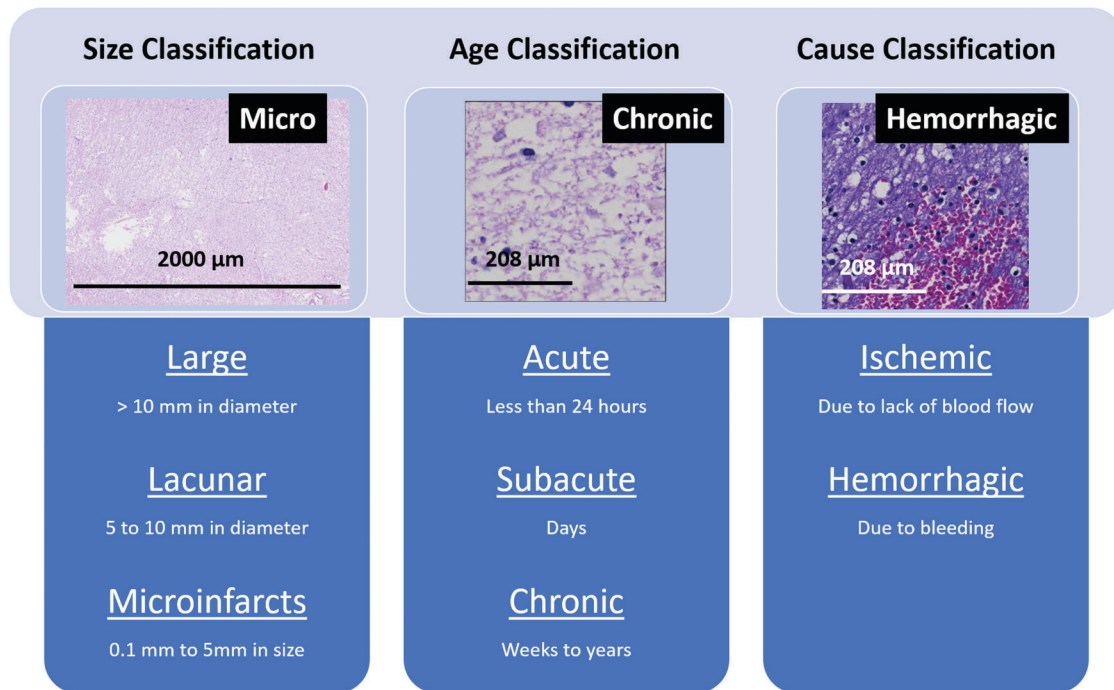


Figure 1. Schematic outlining the different classification categories of infarcts (size, age, and cause), their sub-classifications found in postmortem tissue, and their characteristics. The top row shows examples of lesions for each classification category. For size classification, we show an example of microinfarct on the left. For age classification, we show a chronic infarct, and for cause classification, we show a hemorrhagic infarct. The thresholds for age classification (<24 h for acute, days for subacute, and weeks to years for chronic) were determined according to reference⁴; the thresholds for size classification (over 10 mm diameter for large, 5-10 mm in diameter for lacunar, 0.1-5 mm in diameter for microinfarcts) were determined according to reference.⁵

in imaging modalities such as CT and MRI.^{22–24} To our knowledge, there exist no peer-reviewed publications on automated detection of microinfarcts or microhemorrhages in WSIs of human brains using ML.

To fill this gap, we outline 3 key aspects for the current pipeline, namely to: (1) carefully define the task to be automated, focusing on microinfarcts/microhemorrhages in white matter (WM); (2) create a pilot annotated WSI dataset encompassing a variety of appearances of microinfarcts/microhemorrhages, as well as un-infarcted tissue as controls; and (3) assess the interpretability of the model's predictions on their accuracy in identifying microinfarct/microhemorrhage markers consistent with the physiological definition of the pathologies, and distinguishing them with other pathologies like WM rarefaction (WMR) that may share overlapping features. We primarily focus on convolutional neural network (CNN) models because of their proven efficacy in similar domains,^{18,20,21,25} and their capability for interpretability.²⁶ This article presents an end-to-end, patch-based pipeline for WSI processing and ML model training, aimed at the automated neuropathological analysis of microinfarcts/microhemorrhages in WSIs derived from decedents within the University of California Davis Alzheimer's Disease Research Center (UCD ADRC) brain bank. To our knowledge, this study is the first to automate a workflow using ML for microinfarct/microhemorrhage screening in post-mortem WSIs

derived from human brain, while making the code and images freely available.

METHODS

Cohort

For cohort selection, the UCD ADRC brain bank was queried for cases who had come to autopsy, had a complete neuropathology report utilizing the National Alzheimer's Coordinating Center's (NACC) Neuropathology from version 10 and had available WSIs as of October 2023 (date range was August 2017 through October 2023, $n = 101$). We then determined whether these cases were denoted to have old microinfarcts, old microbleeds, microinfarcts, and/or acute/subacute microhemorrhages/microinfarcts utilizing data collected through the NACC Neuropathology form vs 10 (variables: NPOLD2, NPOLDD2, NPATH4, NPATH6). As the NACC forms do not provide specific regional locations for microinfarcts/microhemorrhages (ie, specific slide/anatomic location), a neuropathology expert (Brittany Dugger [BD]) examined the detailed neuropathology reports of potential cases and corresponding WSIs. The identified locations of microinfarcts/microhemorrhages were then converted from textual data. Given the heterogeneity of brain regions, we selected anatomical areas with the highest prevalence of cases exhibiting microinfarcts/microhemorrhages. Within the UCD ADRC, this was

periventricular WM regions, which have previously been reported to have increase in microinfarcts/microhemorrhages.²⁷

Samples consisted of formalin-fixed, paraffin-embedded 5- to 7- μm sections stained with hematoxylin and eosin (H&E). Of the 101 cases queried, 17 cases had at least 1 microinfarct/microhemorrhage within the frontal, parietal, and/or occipital periventricular WM regions. Furthermore, of the 101 cases, we also selected 5 cases that did not contain any microinfarcts/microhemorrhages in periventricular WM, to serve as an additional control group. Demographics of the 22 cases (17 with and 5 cases without any microinfarcts/microhemorrhages) are listed in [Table 1](#).

All participants or legal representatives approved informed consent during the participants' lives as part of the University of California Davis Alzheimer's Disease Center program. The data collection process followed current laws, regulations, and guidelines (such as sharing de-identified data that does not contain information used to establish the identity of deceased subjects).

Data and annotations

Two scanners were used to generate WSIs. Stained slides from a subset of subjects were digitized by a Leica Aperio AT2 scanner (.svs format) at 0.5 μm per pixel resolution ($N=8$) and at 0.25 μm per pixel ($N=1$). Stained slides of the remaining subjects were digitized by a Zeiss Axio Z1 scanner (.czi format) at 0.25 μm per pixel resolution ($N=13$). The schematic from selected cases to WSIs is shown in [Figure 2](#).

A trained expert (BD) annotated the 66 available slides. We adopted a binary classification system (Yes/No for microinfarct/microhemorrhages - of which we refer to as "infarct" vs "un-infarcted"); only infarcts (regardless of age) having borders within the sampled tissues were traced using ImageScope v12.4.6.7001 (for .svs format WSI) or ZEN 3.0 Blue Edition (for .czi format WSI); cases with infarcts encompassing the entire tissue sample were excluded. Twenty-six WSIs had annotated regions; the remaining 40 were classified as un-infarcted ([Table 1](#)). As seen in [Figure S1](#), tissue affected by WMR can display similar features of infarcts (eg, tissue rarefaction/degradation) as WMR is within the realm of vascular disease.²⁸ Hence, all WSIs were also assessed for a WSI-level WMR semiquantitative score as: 0 (none), 1 (mild), 2 (moderate), and 3 (severe), assigned by BD adapting previously defined criteria.²⁸

Only regions of microinfarcts/microhemorrhages were annotated through tracing and those traces were processed into binary infarct maps (workflow depicted in [Figure 3](#)). Binary maps, matching the dimensions of the original WSI, enabled us to pinpoint the pixels annotated as microinfarcts/microhemorrhages, while remaining blinded to demographic and pathological variables. Another trained expert blinded to demographics/pathological variables, VP, annotated for the presence of microinfarct/microhemorrhages for all 66 WSIs (without delineation). We leveraged those annotations as presence or absence of the pathologies at the WSI level. We grouped the distinct lesions of microinfarcts and microhemor-

rhages (termed infarct in the remainder of this article), as one single category given the small datasets.

Prior to assigning infarct or un-infarcted labels to WSI patches, we leveraged the WSI segmentation model from Lai et al.²¹ to assign background labels (depicted in black in [Figure 3](#)) to the extracted patches. Once background patches were removed and labeled, the remaining patches were labeled according to the binary infarct maps ([Figure 3](#)). A single patch was labeled infarct if more than half of its pixels were annotated as infarct by BD.

The dataset was split into train, validation, and test sets on a case-level split ([Table 1](#)), stratified for infarct ground-truth, WMR ground-truth, and file extension type. The case-level split signifies all WSI from a single case were in a single set with no spillover to a different set. We chose a subject-level split to prevent our ML model from overfitting to the subject-specific features. Due to the class imbalance of infarcts in WSI and subject-level, the subject-level split imposed constraints in our stratification of the split data, as we were not able to maintain identical ratios of infarct samples ([Table S1](#)).

Pipeline

Overview and pre-processing

Ultra-high-resolution WSIs contain a rich set of information, including a wide range of morphological patterns, cellular features, and tissue structures.²⁹ However, using WSIs for ML analysis can be challenging due to their large size, which exceeds the processing capabilities of most commercially available graphical processing units (GPU). This limitation prevents the direct use of segmentation or object detection ML algorithms.³⁰ To address the challenges of using gigapixel WSIs for developing an automated ML system, we used a patch-based method.³⁰ In the patch-based approach, a WSI is initially subdivided into multiple fixed-size patches, such as 512 \times 512 pixels. Each patch is then assigned a class label, such as infarct, un-infarcted, or background. Subsequently, the model is trained using the patch dataset derived from the WSIs and their annotations. The trained model is then used to predict the classification of WSI patches in the validation and test sets. However, the patch-based method does not generalize well, particularly when dealing with dataset heterogeneity, such as varying magnifications of images.³¹ To address this issue, we implemented a pre-processing step harmonizing all WSIs to a consistent resolution of 0.25 μm per pixel prior to both training and evaluation, following methodology of Oliveira et al.³¹ This harmonization process helped ensure the model could effectively handle diverse image magnifications and maintain robust performance across different WSIs.

Model architecture

In this pipeline, we utilized existing deep learning architectures originally developed for general image classification tasks.³² Among these architectures, CNNs have gained notable popularity in the last decade particularly for medical image analysis. CNN deep architectures and significant performance improvements have been noted in many image classification tasks.^{33,34} We used ResNet-18, a residual network commonly used in patch-based ML in the histopathology literature.^{21,35} ResNet-18

Table 1. Case demographics and vascular disease severity.

Case	Race/ ethnicity	Sex	Age (years)	PMI (h)	Vascular brain injury	Infarct patches (frontal region)	Infarct patches (parietal region)	Infarct patches (occipital region)	WMR score (frontal region)	WMR score (parietal region)	WMR score (occipital region)	Split	Group	Scan format
1	White	M	62	7.6	Severe	0	0	1	Mild	Mild	Moderate	Train	Infarcted	SVS
2	White	M	82	110.2	Moderate	0	0	0	Severe	Moderate	Moderate	Train	Un-infarcted	SVS
3	White	M	89	3	Severe	0	0	65	Mild	Moderate	Moderate	Train	Infarcted	SVS
4	White	F	90+	22.7	Severe	0	0	1732	Severe	Mild	Severe	Train	Infarcted	SVS
5	White	M	76	6.2	Moderate	0	0	21	Moderate	Mild	Moderate	Train	Infarcted	SVS
6	White	F	90+	12.2	Mild	145	0	0	Severe	Moderate	Severe	Train	Infarcted	SVS
7	White	M	90+	7.5	Moderate	0	0	1553	Moderate	Moderate	Moderate	Validation	Infarcted	CZI
8	White	F	90+	110.1	Mild	0	0	210	Mild	Mild	Severe	Test	Infarcted	SVS
9	White	M	89	6.1	Severe	1213	0	355	Severe	Severe	Severe	Validation	Infarcted	SVS
10	White	M	90+	4.1	Severe	203	0	107	Severe	Moderate	Moderate	Train	Infarcted	CZI
11	African American	F	89	106.5	Moderate	0	69	3	Severe	Severe	Moderate	Test	Infarcted	CZI
12	White	F	78	110	Moderate	0	0	82	Mild	Moderate	Moderate	Test	Infarcted	CZI
13	White	M	90+	47	Mild	0	57	35	Severe	Moderate	Moderate	Train	Infarcted	CZI
14	White	M	89	4.2	Mild	0	0	0	Moderate	Mild	Moderate	Train	Un-infarcted	CZI
15	Hispanic	F	81	5	Mild	0	0	0	Mild	Mild	Mild	Train	Un-infarcted	CZI
16	White	F	87	11	Mild	0	172	64	Moderate	Moderate	Moderate	Train	Infarcted	CZI
17	African American	F	84	38	Mild	0	0	0	Mild	Mild	Moderate	Train	Un-infarcted	CZI
18	White	F	90+	6.4	Mild	0	23	0	Moderate	Moderate	Moderate	Train	Infarcted	CZI
19	Hispanic	F	90+	6.5	Severe	211	52	6541	Severe	Severe	Severe	Train	Infarcted	CZI
20	Hispanic	F	88	9	Mild	0	0	0	Severe	Mild	Mild	Validation	Un-infarcted	CZI
21	White	F	76	15	Mild	11	96	65	Moderate	Moderate	Moderate	Test	Infarcted	SVS
22	White	M	88	10	Mild	0	17	0	Mild	Mild	Moderate	Train	Infarcted	CZI

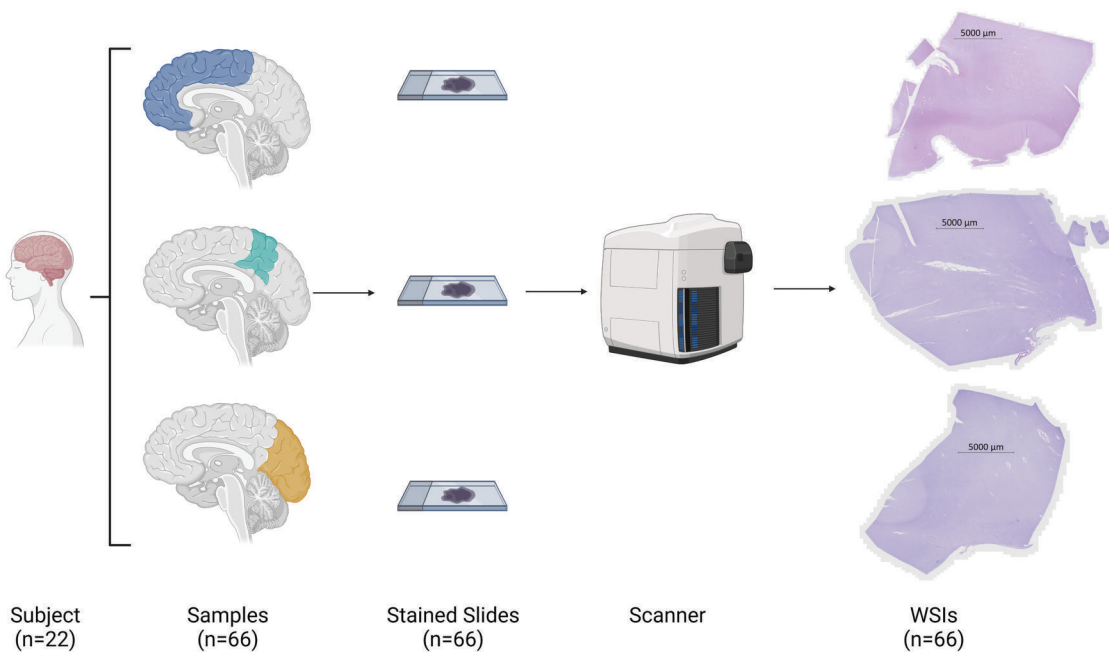


Figure 2. Overview schematic for the regions sampled and whole slide image (WSI) pipeline. Frontal, parietal, and occipital periventricular WM regions were sampled from each of the 22 cases and subjected to H&E staining, then scanned in to create a digital WSI. This figure only depicts the Zeiss scanner; an Aperio scanner was also utilized to generate WSIs as discussed in Methods section. Figure created in BioRender.com. H&E, hematoxylin and eosin; WM, white matter.

incorporates residual connections, which makes networks easier to optimize, leading to improved training.³⁶

Post-processing and detection

After the training phase, the output of our pipeline is a WSI-level prediction heatmap generated using the sliding window approach to systematically process all WSI patches, a technique commonly employed in patch-based pipelines.^{20,21,30} Similar to other ML-based work in pathology,³⁷ we focus on detecting individual pathologies rather than segmenting them. Due to the inherent challenges posed by the imprecise boundaries of infarct lesions,³ we avoided defining our task as a pixelwise segmentation of the infarct region for the current study. Therefore, the heatmap output shown in Figure 3, is an intermediate output of our study prior to generating prediction bounding boxes.

Prior ML work in pathology has shown post-processing steps are able to improve output masks and remove unwanted predictions.²¹ We implemented an additional post-processing step within our patch-based pipeline to refine the prediction areas in the output heatmaps (Figure 3). Initially, we employed the morphological opening operation, a fundamental image processing technique particularly useful for removing small, unwanted noise or fine structures in an image while preserving the overall structure of larger objects.³⁸ In our study, opening operation was utilized to eliminate small false positive predicted regions. Second, we performed the connected component analysis with an area thresholding (AT) to filter out the remaining false positives.³⁹ In this, the same intensity regions of a binary image with areas below the specified threshold are often considered noise or undesired artifacts and are therefore excluded from further analysis.

Lastly, we applied the morphological dilation operation to refine the leftover infarct predicted regions further.⁴⁰

After the post-processing morphological operations were applied, we leveraged the prediction heatmap to define detected infarcted regions. We used distance thresholding (DT) to generate the bounding boxes covering the infarct predicted regions.³⁹ In general, DT is a technique commonly used in image processing to define or limit the spatial extent of certain features or objects within an image based on their proximity to a reference point or region. In the context of generating bounding boxes for infarct predictions, DT ensures the generated bounding boxes accurately encompass the areas identified as infarcts while excluding adjacent regions. An example of the detection output of our pipeline, consisting of a binary output mask with a bounding box denoting the detected infarct region is shown in Figure 3.

Selecting field-of-view

Selecting an appropriate field-of-view (FOV) is an important step in the patch-based analysis of infarct in WSI. Traditionally, pathology-based ML studies have used smaller patch sizes such as 256×256 pixels.^{20,21,41} However, infarct detection can have subjective components, and the affected area may be much larger than other pathologies, thus not fitting the entire lesion in a 256×256 pixels patch. We experimented with 3 distinct patch sizes: 256×256 , 512×512 , and 1024×1024 pixels in the pipeline. We trained, validated, and evaluated separate models for each patch size using the same end-to-end pipeline. Different patch sizes can yield varying degrees of success regarding computational efficiency and the resultant heatmaps used for infarct detection (Figure 4).

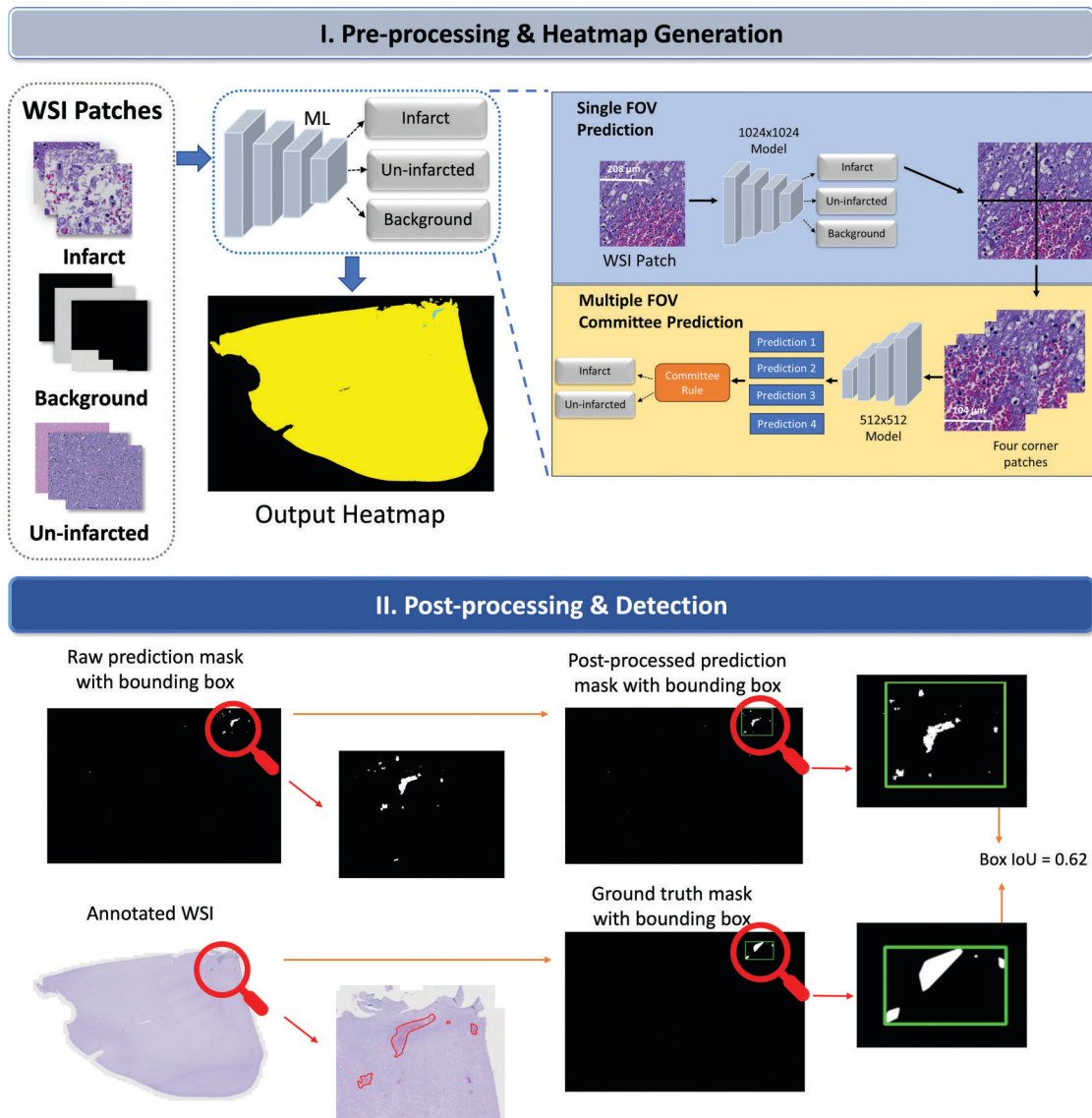


Figure 3. Overview figure for our end-to-end pipeline. The first step for stage I, pre-processing and heatmap generation (light blue box on top), was patch generation. After tiling the WSI, we assigned the patch labels of infarct, un-infarcted, or background. The labeled patches were then used to train the ML model. Once trained, our ML model could output predictions for each WSI patch, which we used to generate heatmaps. In the heatmaps, black represents background, yellow represents un-infarcted tissue, and blue represents infarcted tissue. The first steps of stage II, that is post-processing and detection (dark blue box on bottom), were to take heatmaps and apply morphological closing and opening operations to remove small positive predicted regions from the output. Subsequently, we took the prediction mask and established the infarct prediction bounding box regions to compare their detection accuracy with our ground truth prediction boxes. From the comparison between prediction and ground truth bounding boxes, we acquired the box IoU score metric, which is shown in the top-right portion of this figure. The areas of individual detected infarct regions were later leveraged for WSI-level screening prediction. ML, machine learning; WSI, whole slide image.

To leverage the advantages of small (such as 256×256) and large (such as 1024×1024) FOVs, we proposed a multi-resolution FOV strategy for patch classification depicted in Figure 3. This approach started with evaluating the model using a standard 512×512 or 1024×1024 patch. If the model predicted the presence of an infarct, the patch was further subdivided into 4 256×256 or 512×512 sub-patches,

respectively, each corresponding to a corner of the original patch. The classification of the original larger patch as an infarct was contingent upon a committee decision of the smaller FOV ML model on all 4 of the smaller sub-patches. We evaluated different committee decisions for infarct prediction, such as complete agreement (where all sub-patch predictions must agree on the infarct prediction), partial agreement

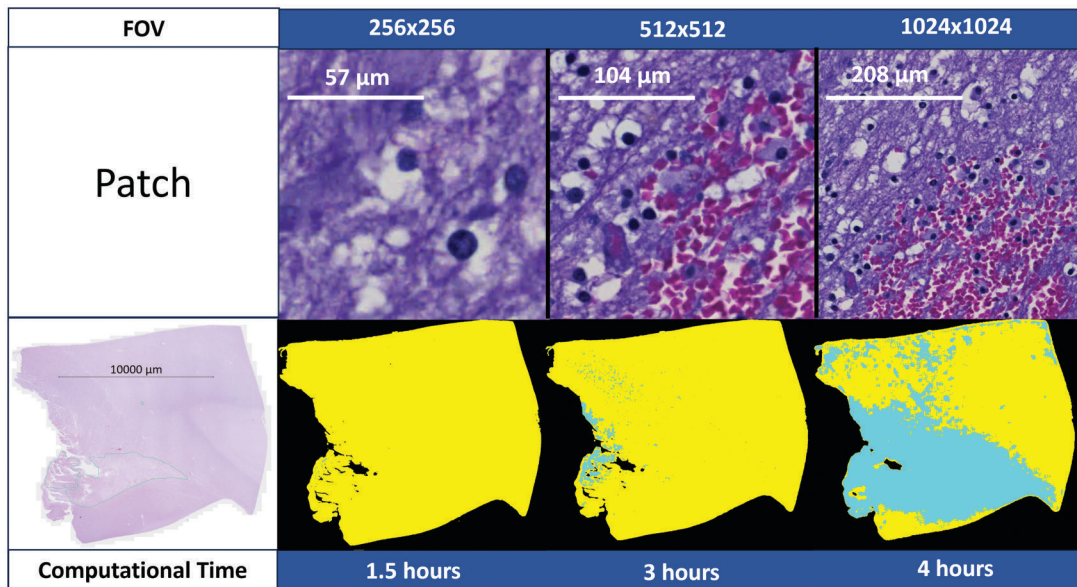


Figure 4. Overview of computational time required for single FOV under different patch sizes. Far left, we have the original whole slide image with green traces denoting BD's infarct annotations. On the right, the heatmap results for different FOVs and their respective generation time. In the heatmap, black represent background; yellow represents un-infarcted; and blue represents infarcted tissue. BD, Brittany Dugger; FOV, field-of-view.

(where up to n sub-patch predictions must agree on the infarct prediction), and confidence threshold (where one patch's infarct prediction confidence must be above a threshold $[t]$). The multi-FOV approach acts as a refined filter augmenting the model's sensitivity to infarcted regions by necessitating a level of agreement among the smaller, more granular classifications.

Assessing model performance

We used the post-processed prediction mask with bounding boxes depicted in Figure 3 to assess the model's detection performance. We followed standard ML definitions for detection,³⁷ that is, the ability to localize the infarcted regions within a WSI. For detection performance assessment, we used the object detection metrics of mean average precision (mAP) and mean average recall (mAR) set forth by the PASCAL VOC challenge,⁴² and utilized by other ML-based detection studies in pathology.³⁷

Given prior work denoting interrater disagreement with respect to infarct delineation,⁶ we could not guarantee detected infarct regions of our model would align with experts other than the one who annotated our dataset. Hence in addition to BD, another expert, VP, provided annotation, (detailed below in the Statistical evaluation section).

Additionally, our model may predict false positives other experts would otherwise consider infarcts. To address the challenge of evaluating our model's performance given the subjectivity of infarct region definition, we also assessed our model's WSI-level screening capabilities. We defined screening as the ability to detect an infarct-positive WSI. Using the validation set for tuning, we defined a diameter threshold for infarct-positive WSI prediction based on individual detection results. If our model predicted any infarcted region on a given WSI

with diameter greater or equal to the threshold, we considered the WSI as infarct positive. If no predictions surpassed the threshold, we considered the WSI as un-infarcted.

Model interpretation

We leveraged Gradient-weighted Class Activation Mapping (Grad-CAM) saliency map²⁶ to interpret the predictions of our CNN models on the WSI patches. Grad-CAM is a technique used in deep learning to aid in understanding and visualizing the decision-making of CNNs. Grad-CAM provides insights into which regions of a given image are most influential in the network's predictions. The saliency maps were generated using the target concept's gradients flowing into the CNN model's final convolutional layer. This approach allowed us to identify features in patches associated with the given prediction.

Statistical evaluation

Cohen's Kappa (K) was used to assess the statistical significance of agreement of WSI-level infarct predictions between the screening model and the BD annotator. Additionally, we used Cohen's Kappa to assess the agreement between BD and VP annotators on a WSI-level (assessing agreement on presence or absence, not localization). Kolmogorov-Smirnov test was used to measure significance in differences in distribution of infarct pixel counts from low and high WMR. All analyses were conducted on test set samples. To conduct analysis, we used Python (Scikit package⁴³) and considered P values of $<.05$ significant.

Implementation details

All 66 WSIs and their annotations are available on Zenodo (DOI: 10.5281/zenodo.13844200). The networks were

trained and evaluated on a system with Nvidia Tesla T4 GPU with 16 GB of VRAM or a Nvidia RTX A4500 with 20 GB of VRAM, and 192 GB of DDR6 RAM. All networks and their respective pre-trained weights were implemented using the PyTorch framework.⁴⁴ For training, we initialized the learning rate of 0.01 and set the batch size number as the largest possible for the given GPU, architecture, and patch size. We reduced the learning rate by a multiplicative factor of 0.1 after every 10 epochs. To optimize the models, we used the Adam optimizer.⁴⁵ The ResNet-18 starting weights are its ImageNet⁴⁶ pre-trained weights. We trained the model for 40 epochs and saved the hyperparameters/weights resulting in the highest balanced accuracy performance on the validation set at the end of each epoch.

We applied data augmentation techniques during training to enhance the ability of the model to generalize across various inputs. This involved randomly flipping the input images both horizontally and vertically given histological images have no canonical orientation (probability of 50% of being applied to each training image). Additionally, all images were normalized to standardize their pixel values before model inference. During training, we oversampled infarct datapoints to address the class imbalance (Table S1). Furthermore, in the evaluation phase of the network, we set a confidence threshold for the patch-wise infarct class prediction. The confidence threshold value was determined by validation set tuning and was set to 99.5% confidence for infarct prediction. Heatmaps were generated with stride 32 for 256 × 256 FOV, and stride 64 for all other FOVs. Additionally, all parameters required for post-processing, multiple FOV evaluation, and WSI-level screening thresholds were tuned on the validation set WSIs to maximize screening performance on validation set (DT = 25 pixels, AT = 25 pixels, kernel size for opening/closing morphological operations = 4 pixels). The reported experiments for multiple FOV used the highest performing committee rule on validation set, which was agreement of 2 or more. All codes and instructions to reproduce results are available on GitHub (<https://github.com/ucdrubinet/Infarct>).

RESULTS

Initial detection performance

After generating the heatmaps, we produced the final output of prediction bounding boxes, where were compared to the ground truth bounding boxes to evaluate detection performance metrics (Figure S2). We observed ResNet-18 trained on 256 × 256 FOV is our best-performing model for this task

with over 0.15 mAP (Figure S2A) and 0.4 mAR for all infarct regions (Figure S2B). If we only considered infarcts with a diameter ≥ 1 mm within the WSI, the performance improved to over 0.35 mAP (Figure S2C) and 0.6 mAR (Figure S2D). The results showed our model performed significantly better for larger infarcted regions, where there are fewer boundary regions (the most challenging regions to predict). Despite competitive mAP and mAR at lower intersection over union (IoU) thresholds, we observed a notable decrease as the IoU threshold increases.

Our model effectively screened infarcted WSI

In addition to evaluating the detection capabilities of our model, we also assessed its screening performance. For screening, we used our model to generate a single label of infarcted or un-infarcted to each WSI. If a given WSI had any infarct ground-truth annotation of any size, we considered it to be an infarcted WSI. Table 2 summarizes screening performance.

Out of all FOVs tested, 256 × 256 FOV-trained ResNet-18 had best performance with 86.67% accuracy on 87.50% sensitivity and 86.67% specificity (Table 2). Our screening evaluation showed our model could identify which WSIs were infarcted without many false positive WSI-level predictions, regardless of the imprecise detection of the infarcted region. In general, larger FOVs (1024 × 1024) performed worse in WSI screening due to their high rate of false positive predictions, which lowered the model's specificity.

The screening and detection results aligned. In both evaluations 256 × 256 FOV-trained ResNet-18 had best performance. This alignment is expected because the screening output is a direct product of the detection output.

Performance improved when leveraging multiple FOVs

In addition to testing single FOV models, we also leveraged single FOV-trained models for multiple FOV evaluation. Figure S3 summarizes our multiple FOV detection experiments, and Table 2 summarizes the screening experiments. We observed leveraging multiple FOVs improved our detection metrics. Our best model was ResNet-18 using both 1024 × 1024 and 512 × 512 FOVs, which displayed over 0.2 mAP (Figure S3A) and 0.4 mAR (Figure S3B) for all infarct regions. If we only considered larger infarcts (diameter ≥ 1 mm), the performance improved to over 0.6 mAP (Figure S3C) and 0.5 mAR (Figure S3D). Both settings showed a modest improvement over the best performing single FOV model. Additionally, we observed an improvement in screening performance when using multiple FOVs. In this setting,

Table 2. Screening performance of the single and multiple FOV model setups tested.

	Single FOV			Multiple FOV	
	1024 × 1024	512 × 512	256 × 256	1024 × 1024 & 512 × 512	512 × 512 & 256 × 256
Accuracy (%)	66.67	86.67	86.67	86.67	100.00
Specificity (%)	57.14	71.43	85.71	100.00	100.00
Sensitivity (%)	75.00	100.00	87.50	75.00	100.00

Performance refers to WSI-level metrics.

our best model was ResNet-18 using both 512×512 and 256×256 FOVs, which successfully screened all test set infarcted WSIs without any false positive predictions on un-infarcted WSI. Table S2 shows the effect of changing the prediction threshold on the WSI screening performance.

Assessing agreement

Taking the WSI-level predictions from our model's screening performance, we assessed its agreement with the WSI-level labels from both annotators, BD and VP. To ensure fair comparison, we only considered test set WSIs ($N=15$), which had not been seen by the model during training. First, we evaluated the agreement between BD and VP, which had an observed an agreement of 0.867 and a Cohen's Kappa of 0.737, which signified significant substantial agreement ($P \ll .05$). We then compared the screening results from the 2 multi-FOV models tested with BD's annotations, (ie, annotations used in training the model). The smaller FOV models (256×256 and 512×512) outputted the same WSI-level predictions as BD (Table 2), which signified Cohen's Kappa of 1, showing perfect agreement ($P \ll .05$). However, the larger FOV models (512×512 and 1024×1024) had lower agreement, with a Cohen's Kappa of 0.867 ($P \ll .05$). In both multi-FOV models, we observed the same level of agreement or higher for annotators BD and VP. The higher agreement with BD showed the model successfully learned from BD's annotations in training.

Saliency map investigation

We further investigated our model's predictions to diagnose the imprecise detection performance observed in our patch-level classification results. To interpret our model's predictions, we used Grad-CAM,²⁶ which allowed us to verify which areas of a given patch contributed the most to its prediction (see methods - model interpretation section). Grad-CAM maps were generated using the 512×512 FOV-trained ResNet-18 on all test set samples. One hundred Grad-CAM maps were manually inspected. A few selected examples of Grad-CAM maps are shown in Figure 5. We observed the model identified features associated with infarcts such as red blood cells outside of blood vessels and macrophages¹ in the true positive prediction. However, in the false positive prediction, the model incorrectly identified a tissue tear and/or WMR as evidence of an infarct.

WMR effect on false positive predictions

We assessed all un-infarcted WSIs ($N=11$) from both test and validation sets to evaluate the number of pixels classified as infarct, grouping these pixel counts based on the severity of their WMR. As there were low numbers in the none/mild groups, we collapsed them with the moderate group to give 2 categories: none to moderate ($N=8$) and compared them to severe ($N=3$). The pixel counts were acquired from heat-maps generated by the 512×512 FOV-trained ResNet-18 (Figure S4). Given these were control slides, all infarct predictions by the model were false positives. Using the Kolmogorov-Smirnov test, we did not find statistically significant differences between the distributions of pixel counts

($P=.84$). However, we observed the average false positive pixel counts in severe WMR were 3 times larger than none-to-moderate WMR slides.

DISCUSSION

In this small pilot study, we provide a proof-of-concept for an integrated workflow to automate microinfarct/microhemorrhage screening in WSIs using ML. By leveraging ML techniques, we propose an open-source ML tool for automated microinfarct/microhemorrhage screening to augment the ability of the experts. While we faced challenges in the precise detection of microinfarct/microhemorrhage regions, our models excelled in screening WSIs for the presence of microinfarct/microhemorrhages, showing promising utility as a digital pathology tool for research.

Our approach notably excelled in screening and achieved high accuracy. The high performance in screening (overall presence/absence of a microinfarct/microhemorrhage on the WSI) is noteworthy, given the heterogeneity of these pathologies and their variable manifestation based on size, age, and hemorrhagic or ischemic nature^{3,5} (Figure 1). Additionally, we observed the model was able to display high levels of WSI-level agreement with the expert it was trained on. We showed the agreement was statistically significant and equal or higher than the observed agreement with other annotators. However, the complexity of the task was evident in the model's struggle to delineate infarct regions precisely, particularly at higher IoU thresholds. This aligns with the inherent difficulties in infarct boundaries (what we term the "fuzzy boundary problem") and classification, as highlighted in the existing literature and evidenced by the variability in expert evaluations.^{3,6} Our manual investigation of saliency maps showed the model may classify infarct patches when features that may be associated with infarcts (including but not limited to tissue rarefaction, macrophages, corpora amylacea, and/or tissue tears), are detected. We hypothesize this occurs due to shortcut learning in training.⁴⁷ This effect increases false positives and may be a challenge in larger FOVs, which include many structures that could be mistakenly linked to infarcts.

Interestingly, implementing multiple FOVs improved the model's detection and screening capabilities, illustrating the potential of ensemble approaches in ML for pathology. We attribute the higher performance of multiple FOVs over single FOV to the reduction of false positive predictions, potentially mitigating effects of shortcut features on prediction through committee prediction. The committee prediction rule requires multiple models to agree on the prediction of a given patch, which reduces the chance of false positive prediction due to learned shortcuts as they may only be present in one of the corner patches.⁴⁷ On the other hand, this same effect leads to a slight reduction of true positive predictions. This explains the slight reduction in mAR on larger infarct detection, which can be attributed to the committee incorrectly filtering out correct infarct predictions.

Existing work addressed classification and segmentation of infarcts in in- vivo imaging,^{22-24,48} but a gap in automated post-mortem histological evaluation of infarcts remains.

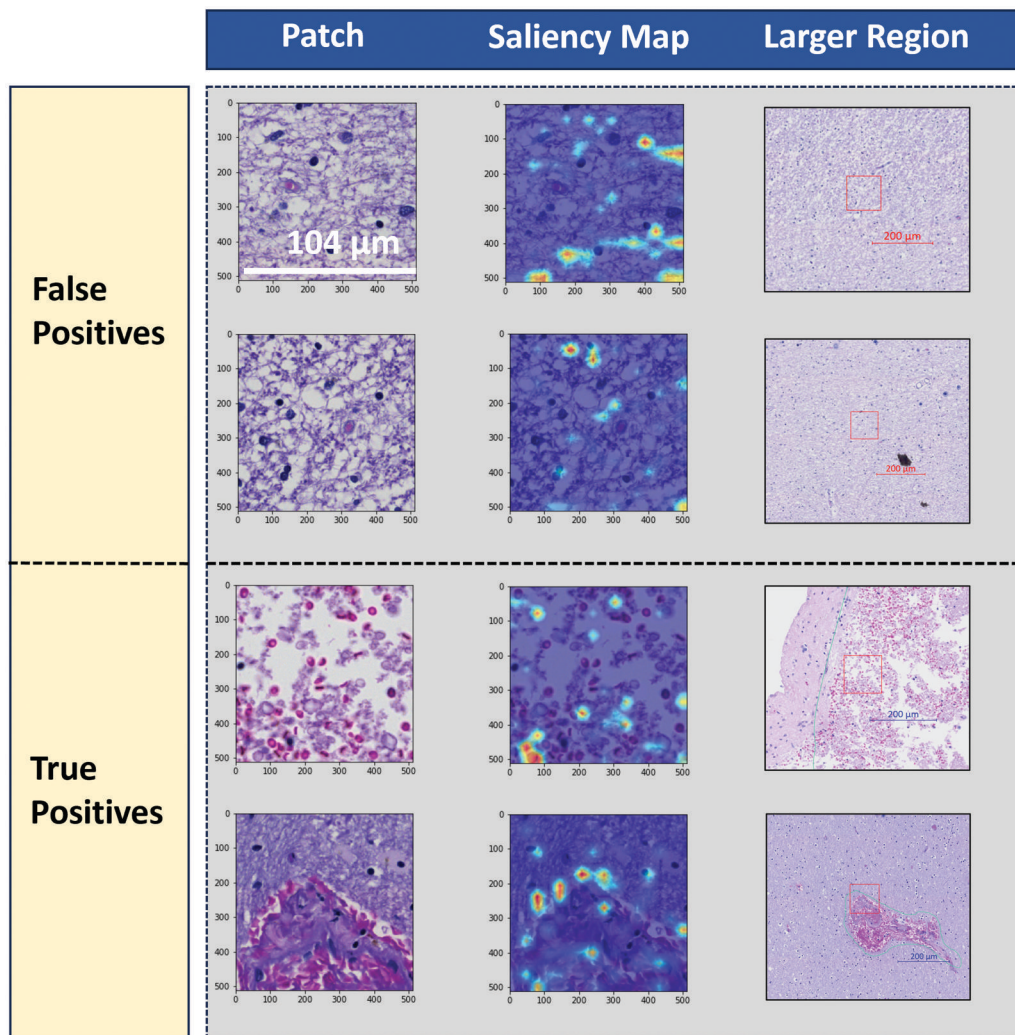


Figure 5. Examples of Grad-CAM saliency maps from the 512×512 FOV of ResNet-18 trained single FOV model. Top panels: examples of false positives (un-infarcted patch incorrectly classified as infarct); bottom panels: examples of true positives (infarct patch correctly predicted as infarct). Far left, we have the original patches (Patch). Middle, are the Grad-CAM (saliency maps) of the corresponding patch. In the Grad-CAM, warmer colors represent higher relevance for the model's predictions. Far right images are macro views (Larger Region) of the 512×512 patch region (red box); green lines are the annotation provided by the expert to denote the infarct boundary; scale bar of $200 \mu\text{m}$. Upon qualitative analysis, for both false positive examples, hot spots are predominately over rarified tissue. For the top true positive example, the hot spots (located at [80, 450] and [200, 330] coordinates) appear to be macrophages; the bottom hot spots (located at [80, 250], [150, 210], [190, 210], and [195, 280] coordinates) line the perimeter of a blood vessel. FOV, field-of-view; Grad-CAM, Gradient-weighted Class Activation Mapping.

Efficiently screening infarcts in WSIs can provide a tool to aid experts, with the potential to enhance deep phenotyping research. Our work contributes to the pathology domain and bridges a gap in digital pathology, providing an initial benchmark to be compared in future works.

Our study is not without limitations. Our data consisted of a small convenience sample from a single institution's Alzheimer's Disease Research Center brain bank focused on a few anatomic areas and one staining method (H&E) for validation of our workflow. Given the noted interrater disagreements in microinfarct/microhemorrhage assessments,⁶ our single-annotator ground truth could be skewed, affecting the learning and generalizability of the model. Furthermore, other pathologies such as WMR posed additional challenges as we lack annotations traces for shortcut learning hypothesis valida-

tion.⁴⁷ Further investigation of WMR effect on false positives is needed as evidenced by the increased false positive rates in higher WMR scores, which was found not statistically significant (Figure S4). Moreover, our Grad-CAM analysis was limited as we manually inspected about 200 patches of the over 400 000 test set patches (Table S1).

Despite limitations, this pilot study aids in filling a critical gap in infarct detection. The novel application of ML in post-mortem automated microinfarct/microhemorrhage screening lays a foundation to aid in new avenues for scalable detection and quantification in future works. The performance metrics of our ML model underscored both the promise and challenges of applying ML in pathology. Furthermore, saliency mapping results aid in model interpretability, affirming the physiological coherence of the predictions of our model

(Figure 5), and show a promising direction for future research addressing shortcut learning in training. By providing an initial workflow and dataset for automated infarct screening, we allow future studies to expand training and evaluation datasets, incorporating WSIs from additional sources, multi-rater annotations in training, breakdown of infarct category into ischemic and hemorrhagic, further exploring stains with focus on vasculature structures, and exploring the integration of additional ML techniques to enhance detection precision. In conclusion, this pilot study advances a model with the potential to provide automated analysis pipelines using ML for augmenting the ability of the expert.

ACKNOWLEDGMENTS

The authors thank the families and participants of the University of California Davis Alzheimer's Disease Research Centers (ADRC) for their generous donations as well as ADRC staff and faculty for their contributions. The authors also thank Zhengfeng Lai for his help reviewing the paper and aiding with machine learning discussions.

SUPPLEMENTARY MATERIAL

Supplementary material is available at academic.oup.com/jnen.

FUNDING

Resources for this study were funded in part by grants from the National Institute on Aging of the National Institutes of Health under Award Numbers R01AG062517 and P30AG072972. This work was funded in part by the HEALHER (Heart, BrEast and BrAin Heath Equity Research) Program supported by residual class settlement funds in the matter of April Krueger v. Wyeth, Inc., Case No. 03-cv-2496 (US District Court, SD of California). A.C.V. is also supported by the Frances Lazda Endowed Chair in Women's Cardiovascular Medicine.

CONFLICTS OF INTEREST

None declared.

REFERENCES

- McAleese KE, Alafuzoff I, Charidimou A, et al. Post-mortem assessment in vascular dementia: advances and aspirations. *BMC Med*. 2016;14:129.
- Whisnant JP, Basford JR, Bernstein EF, et al. Classification of cerebrovascular diseases, III. Special report from the National Institute of Neurological Disorders and Stroke. *Stroke*. 1990;21:637-676.
- Brundel M, de Bresser J, van Dillen JJ, et al. Cerebral microinfarcts: a systematic review of neuropathological studies. *J Cereb Blood Flow Metab*. 2012;32:425-436.
- Birenbaum D, Bancroft LW, Felsberg GJ. Imaging in acute stroke. *West J Emerg Med* 2011;12:67-76.
- Bir SC, Khan MW, Javalkar V, et al. Emerging concepts in vascular dementia: a review. *J Stroke Cerebrovasc Dis*. 2021;30:105864.
- Skrobot OA, Attems J, Esiri M, et al. Vascular cognitive impairment neuropathology guidelines (VCING): the contribution of cerebrovascular pathology to cognitive impairment. *Brain*. 2016; 139:2957-2969.
- Blevins BL, Vinters HV, Love S, et al. Brain arteriosclerosis. *Acta Neuropathol*. 2021;141:1-24.
- Love S, Chalmers K, Ince P, et al. Development, appraisal, validation and implementation of a consensus protocol for the assessment of cerebral amyloid angiopathy in post-mortem brain tissue. *Am J Neurodegener Dis*. 2014;3:19-32.
- Strozyk D, Dickson DW, Lipton RB, et al. Contribution of vascular pathology to the clinical expression of dementia. *Neurobiol Aging*. 2010;31:1710-1720.
- Deramecourt V, Slade JY, Oakley AE, et al. Staging and natural history of cerebrovascular pathology in dementia. *Neurology*. 2012;78: 1043-1050.
- Kalaria RN, Kenny RA, Ballard CG, et al. Towards defining the neuropathological substrates of vascular dementia. *J Neurol Sci*. 2004;226:75-80.
- Barisoni L, Lafata KJ, Hewitt SM, et al. Digital pathology and computational image analysis in nephropathology. *Nat Rev Nephrol*. 2020;16:669-685.
- Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol*. 2019;20:e253-e261.
- Wang S, Yang DM, Rong R, et al. Pathology image analysis using segmentation deep learning algorithms. *Am J Pathol*. 2019;189: 1686-1698.
- Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform*. 2016;7:29.
- McKenzie AT, Marx GA, Koenigsberg D, et al.; PART Working Group. Interpretable deep learning of myelin histopathology in age-related cognitive impairment. *Acta Neuropathol Commun*. 2022;10:131.
- Koga S, Ghayal NB, Dickson DW. Deep learning-based image classification in differentiating tufted astrocytes, astrocytic plaques, and neuritic plaques. *J Neuropathol Exp Neurol*. 2021;80:306-312.
- Vizcarra JC, Pearce TM, Dugger BN, et al. Toward a generalizable machine learning workflow for neurodegenerative disease staging with focus on neurofibrillary tangles. *Acta Neuropathol Commun*. 2023;11:202.
- Marx GA, Koenigsberg DG, McKenzie AT, et al.; PART Working Group. Artificial intelligence-derived neurofibrillary tangle burden is associated with antemortem cognitive impairment. *Acta Neuropathol Commun*. 2022;10:157.
- Tang Z, Chuang KV, DeCarli C, et al. Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nat Commun*. 2019;10:2173.
- Lai Z, Oliveira LC, Guo R, et al. BrainSec: automated brain tissue segmentation pipeline for scalable neuropathological analysis. *IEEE Access*. 2022;10:49064-49079.
- Dogan S, Barua PD, Baygin M, et al. Novel multiple pooling and local phase quantization stable feature extraction techniques for automated classification of brain infarcts. *Biocybernet Biomed Eng*. 2022;42:815-828. <https://doi.org/10.1016/j.bbe.2022.06.004>
- Peng S-J, Chen Y-W, Yang J-Y, et al. Automated cerebral infarct detection on computed tomography images based on deep learning. *Biomedicines*. 2022;10:122.
- Deepa B, Murugappan M, Sumithra MG, et al. Pattern descriptors orientation and map firefly algorithm based brain pathology classification using hybridized machine learning algorithm. *IEEE Access*. 2022;10:3848-3863.
- Li Z, Liu F, Yang W, et al. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans Neural Netw Learn Syst*. 2022;33:6999-7019.
- Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *IEEE*

- International Conference on Computer Vision (ICCV), Venice, Italy. 2017:618-626. <https://doi.org/10.1109/ICCV.2017.74>
27. Arvanitakis Z, Leurgans SE, Barnes LL, et al. Microinfarct pathology, dementia, and cognitive systems. *Stroke*. 1997;42:722-727.
 28. Dugger BN, Clark CM, Serrano G, et al. Neuropathologic heterogeneity does not impair florbetapir-positron emission tomography postmortem correlates. *J Neuropathol Exp Neurol*. 2014;73:72-80.
 29. Higgins C. Applications and challenges of digital pathology and whole slide imaging. *Biotech Histochem*. 2015;90:341-347.
 30. Hou L, Samaras D, Kurc TM, et al. Patch-based convolutional neural network for whole slide tissue image classification. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016;2016:2424-2433.
 31. Oliveira LC, Lai Z, Harvey D, et al. Preanalytic variable effects on segmentation and quantification machine learning algorithms for amyloid-beta analyses on digitized human brain slides. *J Neuropathol Exp Neurol*. 2023;82:212-220.
 32. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:770-778.
 33. Bianchini M, Scarselli F. On the complexity of neural network classifiers: a comparison between shallow and deep architectures. *IEEE Trans Neural Netw Learn Syst*. 2014;25:1553-1565.
 34. Montufar GF, Pascanu R, Cho K, et al. On the number of linear regions of deep neural networks. In: *Neural Information Processing Systems*. 2014;27. <https://api.semanticscholar.org/CorpusID:5941770>.
 35. Hamida AB, Devanne M, Weber J, et al. Deep learning for colon cancer histopathological images analysis. *Comp Biol Med*. 2021;136:104730.
 36. He K, Zhang X, Ren S. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770-778.
 37. Wong DR, Magaki SD, Vinters HV, et al. Learning fast and fine-grained detection of amyloid neuropathologies from coarse-grained expert labels. *Commun Biol*. 2023;6:668.
 38. Soille P. Morphological operators. In: *Computer Vision and Applications*. Elsevier; 2000:483-515.
 39. Chitradevi B, Srimathi P. An overview on image processing techniques. *Int J Innov Res Comput Commun Eng*. 2014;2:6466-6472. <https://api.semanticscholar.org/CorpusID:4838731>
 40. Chanda B. Morphological algorithms for image processing. *IETE Tech Rev*. 2008;25:9-18.
 41. Wang D, Khosla A, Gargeya R, et al. Deep learning for identifying metastatic breast cancer. arXiv, arXiv:160605718, 2016, preprint: not peer reviewed.
 42. Everingham M, Van Gool L, Williams CKI, et al. The PASCAL visual object classes (VOC) challenge. *Int J Comput Vis*. 2010;88:303-338.
 43. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
 44. Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32. <https://api.semanticscholar.org/CorpusID:202786778>.
 45. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv, arXiv:1412.6980, 2014, preprint: not peer reviewed.
 46. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA*. 2009:248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
 47. Geirhos R, Jacobsen J-H, Michaelis C, et al. Shortcut learning in deep neural networks. *Nat Mach Intell*. 2020;2:665-673.
 48. Neethi AS, Niyas S, Kannath SK, et al. Stroke classification from computed tomography scans using 3D convolutional neural network. *Biomed Signal Process Control*. 2022;76:103720.