Pathology-Based Ischemic Stroke Etiology Classification via Clot Composition Guided Multiple Instance Learning

Mara Pleasure*[1, 2] Ekaterina Redekop*[1, 2, 3] Jennifer S Polson[1, 3] Haoyue Zhang[1, 3]
Naoki Kaneko[2] William Speier[1, 2]
Corey W Arnold[1, 2, 3, 4]
[1]Computational Diagnostics Lab, UCLA [2]Department of Radiology, UCLA
[3]Department of Bioengineering, UCLA [4]Department of Pathology, UCLA
{mpleasure, eredekop, jpolson, speier, cwarnold}@ucla.edu
harryzhangbruins@gmail.com nkaneko@mednet.ucla.edu

Abstract

Accurate identification of the etiology of acute ischemic stroke is crucial to prevent secondary strokes. With the growing utilization of endovascular thrombectomy as a treatment for acute ischemic stroke, there has been an increased interest in analyzing the removed clot tissue, opening possibilities for developing histology-based automated diagnostic methods for clot etiology prediction. In this paper, we propose an automated pipeline for Pathology-Based Ischemic Stroke Etiology Classification via Clot Composition Guided Multiple Instance Learning, that leverages the heterogeneity of the main clot components, specifically red blood cells (RBCs) and fibrin, to predict clot etiology solely using digital pathology data. We combine a publicly available dataset from 11 different medical Centers with a private dataset from the University of California, Los Angeles (UCLA). We train the model using center-wise leave-one-out cross-validation to create a model that can generalize to all 12 Centers. Additionally, we compare three different self-supervised methods for embedding the histology data from whole slide images (WSIs) for a downstream task and found that combining two feature sets results in the best performance for our dataset. Our solution resulted in 0.762±0.141 AUC and 0.869±0.139 PRAUC for differentiation between large-artery atherosclerosis and cardioembolic clot etiology. These results hint at the potential to construct a generalizable model for clot etiology prediction. Such a model could assist in treatment planning, thereby helping reduce the likelihood of recurrent strokes.

1. Introduction

In 2019, Stroke was the second leading cause of death worldwide, and based on a study by Pu *et al*., global stroke prevalence is projected to increase by the year 2030 [18, 35]. 87% of strokes are ischemic strokes, caused by a thrombus obstructing blood flow to the brain [38, 10, 27]. Treatment for ischemic stroke aims to restore cerebral blood flow and primarily centers on thrombolysis and endovascular thrombectomy (EVT) [38, 10]. EVT has a high success rate and is optimal for recanalization in patients with large vessel occlusion [10]. The TOAST (Trial of Org 10172 in Acute Stroke Treatment) score currently assigns ischemic stroke etiology using clinical variables and imaging results from CT/MRI, cardiac imaging, and laboratory results following EVT treatment [1]. Understanding clot etiology is critical for treatment planning and prevention of recurrent stroke. For instance, cardioembolic (CE) stroke patients are at higher risk of recurrent stroke compared to patients with large-artery atherosclerosis (LAA) or other etiologies [1].

Secondary prevention for stroke patients varies depending on etiology, highlighting the importance of accurate etiological identification [5]. Typically, patients with a CE stroke will be prescribed anticoagulants, especially if they have concurrent atrial fibrillation, while patients with LAA etiology will be prescribed antiplatelet therapies [5, 25, 16, 37]. The TOAST classification scheme provides a framework for determining stroke etiology but still has drawbacks [39]. Prevalent criticisms are the number of patients classified as undetermined etiology under the TOAST schema or that it is outdated with current CT standards [39]. Following EVT, histopathological analysis of thrombi can be performed to determine clot composition; however, there is currently no standard method for the analysis of clots. To enhance the understanding of how

*Mara Pleasure and Ekaterina Redekop contributed equally.

clot composition impacts patient outcomes, correlating etiology determined by TOAST scores to clot composition would be beneficial, allowing for automated histopathological analysis pipelines to be built. Prior research indicates that thrombi extracted with EVT have a mixture of red blood cells (RBCs), platelets, fibrin, and white blood cells (WBCs) [38]. LAA thrombi have been characterized as having a higher proportion of RBCs than fibrin [7, 22]. In contrast, CE thrombi tend to be enriched in fibrin [7, 22]. However, clot components such as platelets have not shown to be significantly different between LAA and CE clots [22]. For this reason, we choose to focus on using fibrin and RBC content in each clot. An automated approach capable of defining clot composition could act as a corroborative tool for the TOAST score or assist in discerning previously indeterminate etiologies.

Despite the divergent compositions between different etiological categories identified by existing research, there is currently a dearth of automated methods for the analysis of clots removed with EVT. We propose the development of an automated digital pathology and unbiased pipeline that could be used to determine stroke etiology and provide insight into what regions of the thrombi within whole slide images (WSIs) are indicative of clot etiology.

1.1. Prior Work

Removed tissue or clot can be fixed and stained, commonly with hematoxylin & eosin (H&E), Martius Scarlet Blue (MSB), or von Willebrand Factor immunostaining (VWF), for subsequent histopathological analysis. Traditionally, this analysis occurs by a pathologist via microscope using glass slides; however, WSI scanners have allowed digitized tissue sections at gigapixel resolution. Leveraging WSIs, computer-assisted diagnostic (CAD) platforms can be developed for various tasks, including ischemic clot etiology prediction. WSIs are large high-resolution images with rich detail for mining textural features. However, due to their size, WSIs cannot be directly inputted into a machine-learning model, necessitating feature extraction or patching techniques.

Brinjikji *et al*. used a computer software system named Orbit to quantify the RBC and fibrin content in 1350 clot WSIs stained with H&E and MSB [8]. Using an ensemble of classifiers, including logistic regression, gradient-boosted trees, neural networks, and random forests to classify stroke etiology based on these quantified measurements, they attain an area under the receiver operating characteristic (AUROC) of 0.55 [8]. Patel *et al*. achieved higher performance for predicting clot etiology by extracting 227 engineered histomic and WBC nuclear features from 53 H&E-stained clot WSIs, reaching an AUROC of 0.87; however, this dataset is small and from only one Center [33]. Since clots extracted via EVT are still not routinely analyzed by histopathologists, we believe there could be a benefit in using deep features instead of hand-crafted features for predicting etiology from clot WSIs. Deep features require less prior knowledge and manual preprocessing, and since it is currently unknown precisely what signal is useful in the WSI for predicting etiology, it can be helpful to train a deep learning model to discern the salient areas.

Due to the large size of WSIs, methods such as down-sampling and patching are employed to facilitate the training of deep-learning models. Down-sampling, however, can reduce the information in the image significantly, making it a suboptimal method for this data format. On the other hand, patching preserves the richness of information in the WSI image while making it a manageable size for a deep learning model. When patching, the assignment of labels needs to be thoughtfully considered based on the task's goals. Annotations for regions of interest in WSIs are expensive and time-consuming to collect, meaning these images usually have only slide or patient-level labels. Upon tiling, tiles can be assigned the slide-level label, even if that tile might not have any information indicative of that label. An alternate method for handling tile labels from WSIs is Multiple Instance Learning (MIL), which uses the global label as a weak local label by assigning the patches to a bag with a bag-level label [9]. This can be helpful when pixel-level annotations are unavailable because it can be unclear what patches indicate the slide-level label, resulting in noisy labels if each patch is assigned the slide-level label [31]. In the MIL framework, images are first patched into smaller tiles and collected into a 'bag' where each 'bag' is assigned the slide level label. This 'bag' is fed through a deep-learning model, and each tile is assigned a score. Attention-MIL, a method developed by Ilse *et al*., implements a neural network method to pool the tile-level scores into a bag-level score resulting in a prediction for the whole slide [23]. This method can be interpreted as an attention mechanism that provides attention scores for each patch corresponding to the relative value that the patch had in the bag-level score. The size of the bags can pose memory-related challenges. To circumvent these issues, strategies aimed at reducing dimensionality, such as sampling or embedding, can be used to fit the entire bag within memory. We believe deep features could be helpful in stroke etiology prediction from WSIs since they require little prior knowledge about what types of structural features may provide meaningful insights.

Training a deep feature embedding module with a MIL module end-to-end is possible. However, sufficient data is needed to ensure this feature embedding module can capture informative features from each tile. Alternatively, frozen pre-trained models can extract deep features to feed through the MIL module. Usually, these pre-trained models can be supervised ResNet models trained on out-of-

domain data such as ImageNet [20, 31]. However, relying on pre-trained models using supervised tasks can result in an embedding space that is less generalizable to diverse downstream tasks since the model is trained to extract features for a specific task [32]. Additionally, obtaining pixel-level labels for large histopathology datasets can be expensive due to the expertise required, unsupervised pre-training methods can circumvent this issue. Inspired by natural language processing methods, unsupervised pre-training is an alternative method that can garner general features using methods such as contrastive learning on large unlabeled datasets [14]. The primary objective of unsupervised contrastive learning models is to learn an embedding to encode the shared information between the higher dimensional signal in the data while reducing the dimensionality and discarding less informative signals [32]. Contrastive learning combined with MIL was successfully applied by Li *et al*., who pre-trained SimCLR, (a simple framework for contrastive learning of visual representations) on the Camelyon16 dataset to create an unsupervised model to embed patches for a dual-stream MIL model for the prediction of cancer on a WSI [28].

In this paper, we propose a novel contrastive learning pre-training pipeline with RBC and fibrin guidance for WSI classification of stroke etiology. Using clots extracted with EVT and stained with MSB, we compare and combine three different contrastive learning methods – SimCLR, Momentum Contrast (MoCo), and knowledge distillation with no labels (DINO) to extract embeddings for each bag of patches which are then concatenated. Additionally, we incorporate a guidance module using RBC and fibrin masks. Our goal is to aid the model in prioritizing particularly informative patches for the specific etiology. These are then processed through an Attention-MIL model. The main contributions of this paper are:

- Development of a novel CAD pipeline that leverages RBC and fibrin guidance to predict clot etiology using WSIs.
- A comprehensive comparison of contrastive learning methods for extraction of informative features from clot WSIs
- A training implementation to create a model that can generalize to data from 12 different medical Centers.

2. Materials and Methods

Mayo Clinic organized the Stroke Thromboembolism Registry of Imaging and Pathology (STRIP) AI challenge to find an algorithmic solution to classify ischemic stroke etiology using thrombi WSIs [3]. WSIs were collected via a large multicenter project led by Mayo Clinic Neurovascular Lab. Each WSI was assigned one of the two major ischemic stroke etiology subtypes: CE and LAA. The assignment was performed at each center individually using the TOAST criterion, which does not use pathology data [8]. Various approaches were developed to solve the task and presented during the challenge, but many of them struggled from time and memory constraints due to a challenge format. In this work, we develop a novel framework, which we compare to the solution with the best result in the challenge (see Section 2.4) on the private dataset from the Kaggle leaderboard.

2.1. Dataset

In this work, we use the public STRIP AI Challenge dataset in combination with a private dataset from UCLA to automatically predict clot etiology [8]. We use digitized pathology slides of the thrombotic material extracted mechanically during acute neurovascular procedures. We also use self-supervised learning to create a rich embedding of our WSIs. The contrastive learning methods used in this paper for self-supervised training operate on patches and require a large number to learn deep histological features before they can be effectively applied to classification tasks. To make sure there is no information leakage from the self-supervised model, it must be pre-trained either on the training subset or a separate histological dataset. 12-fold cross-validation experiment setting used in this study (see Section 2.3) makes it challenging to pre-train 12 separate models for 12 different training subsets. Therefore, we used a separate dataset of histological scans of lymph nodes (Camelyon16) to pre-train the three contrastive learning models.

2.1.1 Mayo Clinic - STRIP AI challenge Dataset

The STRIP dataset comprises 754 whole slide images (WSIs) from 632 patients. Clot material was initially collected from 11 different medical Centers and then shipped to a central core laboratory for standard tissue processing, including MSB staining and scanning [8]. The dataset is publicly available from the challenge website [8, 3]. The clot etiologies were categorized as either CE or LAA using TOAST criteria and self-reported from each Center [8]. The center-wise distribution of the number of samples from each category is presented in Table 1.

2.1.2 UCLA Dataset

The dataset consists of 19 MSB-stained WSIs collected from 19 patients and scanned at 40x magnification at UCLA. The distribution of etiologies in this private dataset are 16 CE and 3 LAA cases. The UCLA dataset was used as a 12th Center in combination with the challenge dataset.

Center	1	2	3	4	5	6	7	8	9	10	11	UCLA
CE	44	26	22	88	29	24	70	14	14	37	179	16
LAA	10	3	27	26	9	14	29	2	2	7	78	3

Table 1. Per-Center distribution of CE and LAA clots.

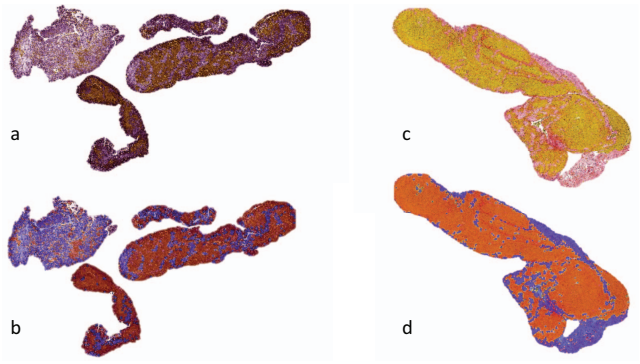

Figure 1. Example of the blood clot WSI from the STRIP AI challenge dataset. a) original WSI thumbnail from Center 11; b) overlay thumbnail with RBC mask (red) and fibrin mask (blue). c) Original WSI thumbnail from Center 4; d) overlay of RBC mask (red), and fibrin mask (blue).

2.1.3 Camelyeon16 Dataset

The data from the Camelyon16 challenge contains 400 whole-slide images (WSIs) of sentinel lymph nodes. The data is stained with H&E and scanned at 40x magnification in two different medical Centers and is publicly available via the challenge website [6].

2.1.4 Preprocessing

First, WSIs were preprocessed to remove the background using thresholding of the thumbnail extracted at the lowest magnification stored in the .tiff pyramid to produce tissue masks. Tissue masks were then refined by applying various morphological operations to fill small gaps and remove noise. As a single magnification level was available for all slides from the STRIP AI challenge dataset, this unknown magnification was used to extract patches of 256x256 from the grid with 0 overlap. Patches were discarded from the analysis if they contained less than 50% of the tissue.

To obtain patch-based RBC and fibrin content values for training our proposed framework, each patch was transformed into HSV color space, and a threshold was applied using the hue channels. Morphological closing was used to fill small gaps and remove noise. To ensure consistent thresholding results in the presence of staining variations across different Centers, the Reinhard stain normalization technique was applied to every patch before finding the optimal threshold [36]. An example of the segmentation result is shown in Figure 1.

2.2. Methods

2.2.1 Our Pipeline

As discussed in Section 1.1, training neural networks directly using WSIs is not feasible due to their size, and therefore slides are usually divided into smaller patches. Many baseline solutions are patch-based neural networks that employ patch-based label assignments and use mean average pooling across patch predictions during evaluation. However, the effectiveness of these approaches can be limited since only specific regions of the input image are typically informative for the slide-level label. Therefore, we formulated the clot etiology classification problem using the attention-based MIL framework [23]. The whole image is a set (or 'bag') of instances with a single classification label corresponding to the entire bag. The relative contribution of instances for the final prediction is then learned through a trainable attention module.

Because many instances are extracted from a single image, using raw image patches to perform classification is computationally intensive. Instead, we used pre-trained models as encoders to extract a low-dimensional set of features for every bag of patches in the dataset (see Section 2.2.2). A fully connected neural network is then used to tune these features to stroke-histology-specific representations, which serve as input to the attention module. As the composition of the clot could provide valuable insight into the underlying etiology, we develop a novel approach that introduces prior knowledge about the clot composition to the attention module. The overview of the proposed framework is shown in Figure 2. The feature representations obtained from transfer learning underwent two distinct, fully connected neural networks. One network was responsible for fine-tuning the features toward RBC representations, while the other focused on fibrin representations. After the fully connected layers, separate attention modules were applied. The output attention vector was subsequently combined with another vector, where each value represented the percentage of the patch occupied by either RBC or fibrin.

2.2.2 Contrastive Learning

Self-supervised learning was used to extract low-dimensional discriminative features from raw image patches by learning from a large scale of unlabeled data. Contrastive-based self-supervised learning approaches are commonly used without the limitation of model architecture [2]. In this work, three different frameworks

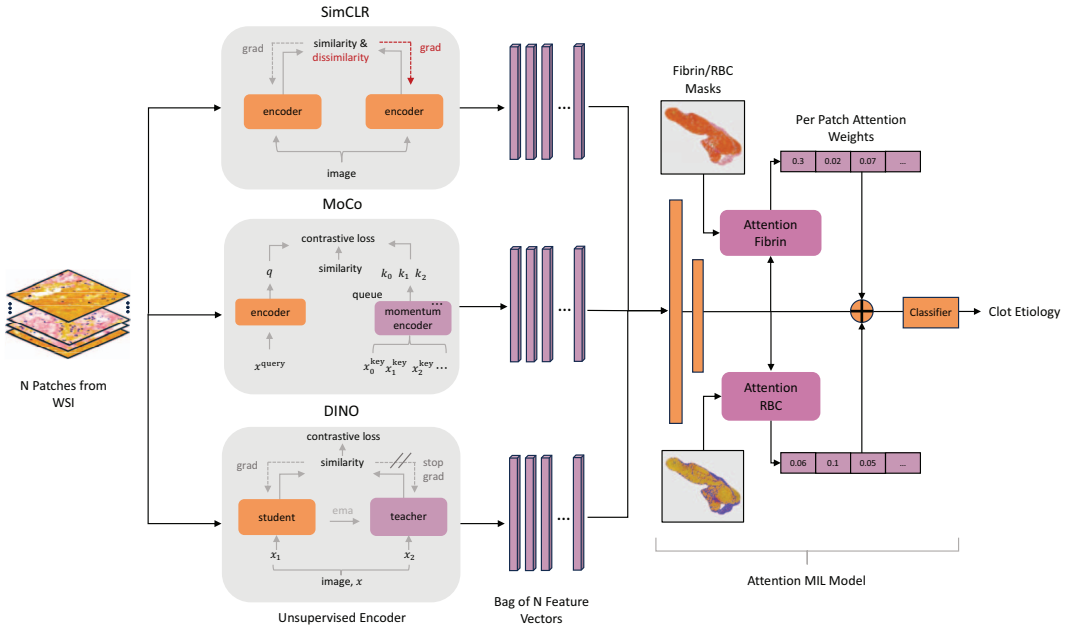

Figure 2. Overview of our proposed framework. The model consists of two separate stages: initial pre-training of the contrastive learning framework on a separate digital pathology dataset, and then training of the attention-based multiple instance framework with biological guidance.

were compared to study the application of contrastive learning to the novel field of stroke pathology and find the approach that leads to the best performance in stroke etiology classification tasks. The SimCLR approach uses various image transformations like random cropping, random flipping, color jittering, and others to form positive and negative pairs. Negative pairs were formed by applying transformations to different patches within one batch [13]. The quality of the resulting representations highly depends on the number of negative pairs used during training, which requires larger batch sizes and, therefore, larger computational resources. MoCo was developed to provide an alternative approach for generating negative samples. It introduces a momentum design to maintain a queue of negative samples, so they are not limited to views from the same batch enabling a smaller training batch size [21]. Vision transformers [17] were adapted to contrastive pre-training after they showed improved performance in supervised vision tasks compared to CNNs. DINO is a framework built based on the idea that aligning positive pairs alone is sufficient to perform well in self-supervised visual learning [12].

2.3. Implementation Details

Site-to-site variability in the interpretation and implementation of the TOAST criteria may result in noise in the training labels [8]. To account for the noise, we utilized a center-wise cross-validation strategy to find the best model setting that will be well-generalizable across different medical Centers. In this setting, 11 Centers were used for training, while the 12th Center was held for validation.

The Camelyon16 dataset was the basis for contrastive pre-training. This dataset has gained significant popularity across various applications and has demonstrated strong performance in numerous downstream tasks [24, 26, 15]. SimCLR requires large computational resources to achieve the best performance in learning representations. In this study, we leveraged the pre-trained weights of the SimCLR model trained on the publicly available Camelyon16 dataset [28]. We pre-trained the MoCo and DINO models on Camelyon16 for two weeks on 4 Quadro RTS 800s GPUs and 4 Tesla V100-SXM2 GPUs, respectively.

To account for label imbalance (see Table 1), we used focal loss to address the class imbalance and mitigate the potential of the model being biased toward the dominant class [29]. The parameters of the loss function were fixed for all experiments alpha set to 0.8 and gamma set to 2.0.

Three different contrastive learning techniques were pre-trained in a self-supervised manner using the Camelyon16 dataset. The pre-trained models were then used to generate deep representations for patches extracted from clot WSIs. Due to the differences in the architectural design of underlying DNNs, the sizes of the representations varied depending on the utilized method. SimCLR, MoCo, and DINO resulted in 512, 1000, and 384 feature vectors, respectively. In this work, we compare the performance of the classifica-

tion model trained based on either individual feature vectors or various combinations of them using concatenation. The initial learning rate was set at 1e-3 and was decreased using the cosine annealing scheduling strategy. We used the Adam optimizer and a batch size of one.

2.4. Comparison Methods

Since this dataset originates from Kaggle, we wanted to compare our method to the winning solution on the leaderboard. The winning solution is an ensemble model based on CoaT [40] and Swin-Large [30] models that assign the slide level label to each patch. The preprocessing steps included subsampling each slide's total number of patches to the top 16 darkest patches. To compare the top leaderboard model to ours, we trained the CoaT and SWIN-Large models with the same experimental setup as ours regarding the 12 folds hold-out center cross-validation. The training code and model weights for this model were not provided in the Kaggle submission, so we utilized the resources publicly released by the authors of the original Swin implementation [30]. Additionally, a pre-training step with MoCo contrastive learning was also specified in the winning solution's submission; we pre-trained on Camelyon16 to provide a balanced comparison.

2.5. Evaluation metrics

As the hold-out set in every fold is imbalanced towards the 'CE' etiology class (see Table 1) except for Center 3, we utilized metrics that are not biased towards the majority class, specifically AUROC and AP computed from ROC and precision and recall (PR) curves, sensitivity, and specificity.

3. Results

3.1. Comparison of Contrastive Learning Methods

Table 2 shows AUROC, Sensitivity, Specificity, and PRAUC across all 12 folds with standard deviation. The model that combined MoCo + DINO features had the highest AUROC of 0.762 and the highest specificity of 0.757. The combination of MoCo + SimCLR features had the highest PRAUC of 0.871. There was no clear standout across the models, however, when looking at all four metrics, MoCo + DINO scored the highest in AUROC and specificity, so it was chosen for the remaining experiments. Surprisingly, the combination of all three feature sets resulted in the lowest performance in terms of AUROC and specificity but had a high sensitivity.

Table 3 breaks down the performance of our model trained on MoCo + DINO features per Center. The model performed the best on the UCLA dataset with an AUROC of 1.0, with the next best performance on Center 9 with an AUROC of 0.928. Overall, the model performed very well on 7 out of the 12 Centers with an AUROC in the 0.7 or higher range, and PRAUC in the 0.8 or higher range. The Center with the lowest performance, Center 3, is the only Center with more LAA to CE cases. It is possible that the model did not generalize to a different label distribution and had a poorer AUROC performance due to this difference. This wide range in performance explains the large standard deviation values for each testing metric.

3.2. Ablation study and Results of Comparison Methods

Given the prior findings of LAA clots having higher RBC content and CE clots having higher fibrin content (see Section 1.1), we sought to measure model performance without fibrin and RBC attention guidance. We ran the same feature combination experiments with feature sets from SimCLR, MoCo, and DINO; however, we used a simpler single attention module that has no prior guidance for RBC or fibrin. The best performance in our ablation study was an AUROC 0.739 (0.139) and a PRAUC of 0.86 (0.138) for the MoCo + DINO feature set model. The addition of a targeted attention module seems to help our full model attend to relevant regions in the WSI that could be informative of the label instead of relying on the model training attentions with no guidance from fibrin or RBC. The top leaderboard model on Kaggle did not perform as well as our models on this training setup, achieving an AUROC of 0.605 (0.139) and a PRAUC of (0.618, 0.0613). It is possible the model had issues with generalizing to unseen data; however, a complete comparison cannot be done since we do not have access to the exact training code.

4. Discussion

In this paper, we developed a novel framework that was trained on clots removed by endovascular thrombectomy (EVT) from 12 different medical Centers. One of the challenges of this task is that ground truth etiology labels were determined at each site by the TOAST score method, which does not utilize stroke histology [8]. As no documented clinical standard for pathological features predicts clot etiology, the interpretability of the model performance is an essential part of our analysis. Figure 3 shows the per-patch attention weights for the BRAISE model and the ablation model trained on MoCo + DINO features. The redder the patch, the more the model used that patch to inform its prediction. There is a noticeable difference in how the attention is localized between the two models. Earlier studies have demonstrated that LAA clots exhibit higher levels of textural heterogeneity, characterized by a predominance of RBC [33]. A statistically significant difference in RBC density between LAA and CE clots was also shown in the study performed by Brinjikji *et al*. [8]. As shown in Figure 3, our model has high attention on RBC regions in both the CE

Pre-training Method	AUROC (std)	Sensitivity (std)	Specificity (std)	PRAUC(std)
MoCo	0.715 (0.130)	0.853 (0.171)	0.556 (0.237)	0.867 (0.117)
SimCLR	0.694 (0.149)	0.799 (0.165)	0.626 (0.220)	0.851 (0.113)
DINO	0.679 (0.095)	0.768 (0.212)	0.621 (0.185)	0.840 (0.149)
MoCo + SimCLR	0.716 (0.103)	0.748 (0.239	0.701 (0.111)	**0.871 (0.121)**
DINO + SimCLR	0.706 (0.112)	0.750 (0.220)	0.659 (0.109)	0.870 (0.092
MoCo + DINO	**0.762 (0.141)**	0.745 (0.220)	**0.757 (0.119)**	0.869 (0.139)
MoCo + DINO + SimCLR	0.646 (0.105)	**0.892 (0.138)**	0.412 (0.225)	0.845 (0.118)

Table 2. Comparison of three contrastive learning methods across our four-test metrics, AUROC, Sensitivity, Specificity, and PRAUC.

Center	1	2	3	4	5	6	7	8	9	10	11	UCLA
AUROC	0.805	0.884	0.554	0.587	0.724	0.723	0.691	0.923	0.928	0.737	0.590	1.0
Sensitivity	0.8	1.0	0.653	0.423	0.555	0.428	0.724	1.0	1.0	0.857	0.506	1.0
Specificity	0.75	0.807	0.545	0.761	0.827	0.916	0.671	0.846	0.785	0.567	0.6666	0.937
PRAUC	0.950	0.986	0.492	0.821	0.854	0.808	0.854	0.989	0.989	0.934	0.754	1.0

Table 3. Comparison of performance of MoCo + DINO model across the four test metrics AUROC, Sensitivity, Specificity, and PRAUC for each Center.

and LAA clots. The model may focus on the amount and pattern of the RBCs in the clots to differentiate between CE and LAA clots. Both CE and LAA clots should have RBCs, but the amount and arrangement with fibrin could be informative to the model. The ablation model has a more scattered signal; for example, in the CE thrombus, the attention map shows a mixture of red patches in the top right tissue. In contrast, our model localizes on a few clot pieces with higher RBCs. Surprisingly, our model localizes to RBCs not only on LAA thrombi but also on CE clots.

Since the data spanned 12 different Centers, we observed a stark difference between many slides regarding color saturation due to staining protocol and scanning. We applied TorchStain [4] to perform stain normalization as a preprocessing step but failed to observe any performance improvement so we omitted it in our final model pipeline.

One of the goals of this work was to develop a solution that would be generalizable across different Centers. A Center-based cross-validation training strategy was utilized to account for differences in the data provided by different Centers. However, different medical Centers could have widely varying proportions of CE to LAA clots, as observed in Center 3. Overall, there was a strong imbalance in the data of CE to LAA. One method we used to address this was weighted sampling. Based on the final pipeline results, we believe that training a model with one Center hold-out cross-validation and ensembling can allow for a model that generalizes to unseen centers, an approach we plan to further validate on new unseen Centers.

Analyzing per-Center metric values, we noticed a high variance of AUROC. We assume that some Centers may perform worse due to inter-site variability in the interpretation and implementation of the TOAST. Noisy labels in the dataset could potentially explain the performance drop for Centers 3 and 4. We believe training our pipeline on a large dataset from one Center may help support this statement.

One of the limitations of the public dataset is that no magnification or resolution information is provided for the STRIP AI challenge dataset. Magnification information is particularly important when a separate dataset is used to pre-train a self-supervised learning framework. Learned deep representations can potentially contain resolution-dependent tissue properties. Consequently, the features extracted from the main dataset using pre-trained encoders may be inaccurate, leading to overall lower performance. The Camelyon16 dataset extracted at 20x magnification was used to pre-train all semi-supervised learning frameworks. The UCLA dataset was originally scanned at 40x magnification; however, the manual assessment revealed that using a 5x resolution is necessary to align with the quality of the STRIP AI challenge dataset. Another limitation of this dataset is a selection bias to clots that are more amenable to removal with EVT, and our dataset could potentially be more imbalanced towards CE due to this selection bias. One method to address this would be to pull more data from different Centers and to investigate if the larger number of CE clots in the dataset is due to population patterns or just a data selection bias in this dataset. Another limitation of the dataset is that it only focuses on clots labeled as CE and LAA, however, there are other possible clot etiologies such as cryptogenic. Future implementations of our pipeline could be trained on datasets with additional clot etiologies to improve the relevancy of our pipeline to other populations.

The limitations of our method primarily stem from the assumptions inherent to our MIL model. When patching WSIs into 256 x 256 images, we risk overlooking the larger spatial information of the WSI and could potentially be losing important architectural features of the clot [31]. In the future, we could incorporate a positional encoding module

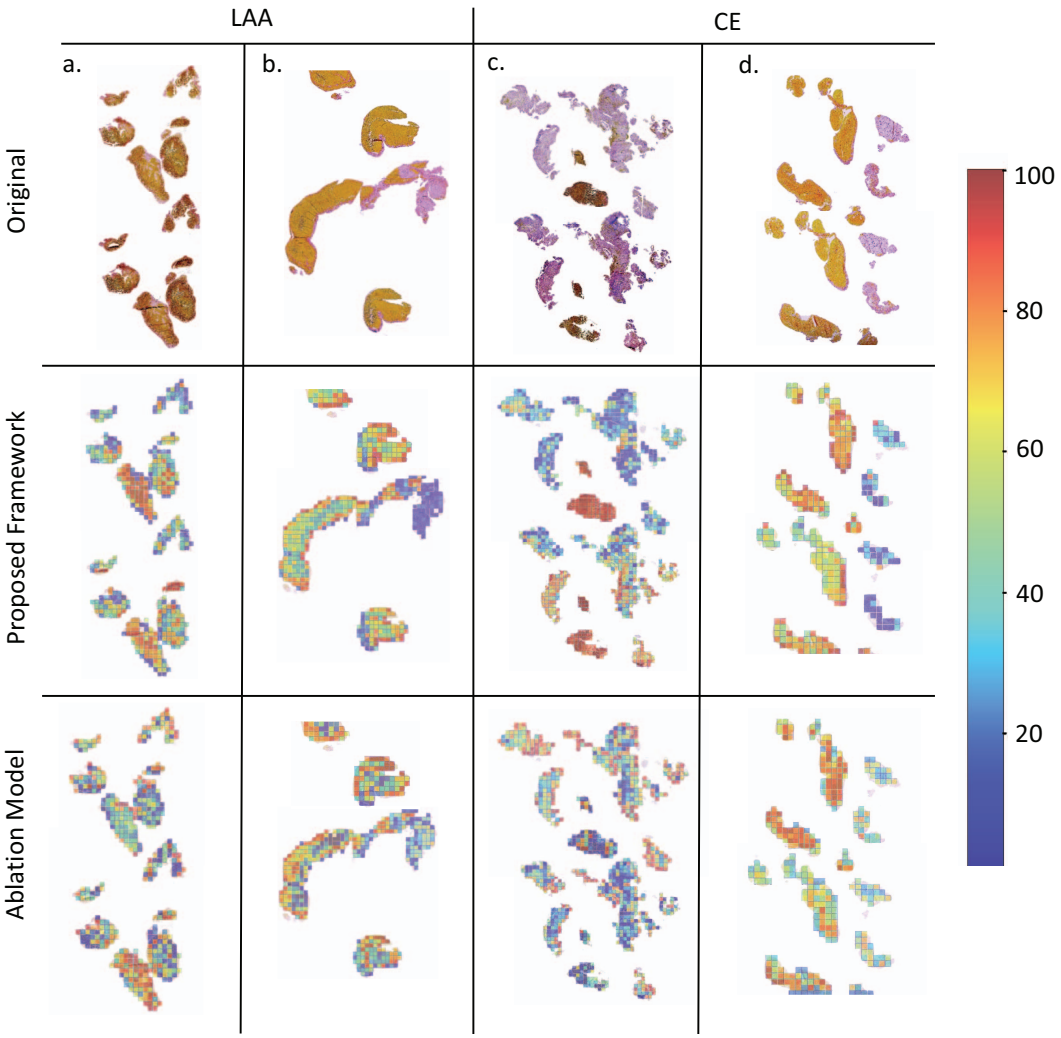

Figure 3. WSIs overlaid with attention maps generated from the attention module of our pipeline for WSIs from Center 11 and Center 7. Attention weights learned by RBC and fibrin branches were averaged to get the final attention value for each patch. The second row shows the results of our final model; the third row shows the results of the ablation model trained without biological guidance. The first two columns correspond to WSIs having LAA stroke etiology from Centers 11 and 7, while the last two columns correspond to WSIs having CE etiology both from Centers 11.

into our pipeline to record patch relationships. Memory-saving methods for WSIs such as streaming convolutional methods that use gradient checkpointing [34] should be considered in the future step to enable a larger look at the clot WSI without the need for downsampling or patching. Another limitation of MIL models, in general, is the assumption that patches are independent and identically distributed [19, 11]. In reality, patches from the same WSI often share certain characteristics, such as contrast and staining, suggesting a degree of correlation and similarity [19, 11].

5. Conclusions

This work is the first to develop a fibrin/RBC guided histology-based blood clot etiology prediction method. By combining contrastive learning and fibrin/RBC guidance we have developed a model that can generalize to unseen data and localize attentions to regions that are predictive in the clot WSIs. Through comprehensive validation, the developed solution has the potential to revolutionize personalized stroke care and enhance secondary prevention strategies.

References

[1] Harold P Adams Jr, Birgitte H Bendixen, L Jaap Kappelle, Jose Biller, Betsy B Love, David Lee Gordon, and

EE Marsh 3rd. Classification of subtype of acute ischemic stroke. definitions for use in a multicenter clinical trial. toast. trial of org 10172 in acute stroke treatment. *stroke*, 24(1):35–41, 1993.

[2] Saleh Albelwi. Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging. *Entropy*, 24(4):551, 2022.

[3] Ryan Holbrook Sobhi Jabal VikashGupta Ashley Chow, Barbaros. Mayo clinic - strip ai, 2022.

[4] Carlo Alberto Barbano and André Pedersen. Eidoslab/torchstain: v1.2.0-stable, Aug. 2022.

[5] Derek Barthels and Hiranmoy Das. Current advances in ischemic stroke research and therapies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1866(4):165260, 2020.

[6] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

[7] Tobias Boeckh-Behrens, Justus F Kleine, Claus Zimmer, Frauke Neff, Fabian Scheipl, Jaroslav Pelisek, Lucas Schirmer, Kim Nguyen, Deniz Karatas, and Holger Poppert. Thrombus histology suggests cardioembolic cause in cryptogenic stroke. *Stroke*, 47(7):1864–1871, 2016.

[8] Waleed Brinjikji, Raul G Nogueira, Peter Kvamme, Kennith F Layton, Josser E Delgado Almandoz, Ricardo A Hanel, Vitor Mendes Pereira, Mohammed A Almekhlafi, Albert J Yoo, Babak S Jahromi, et al. Association between clot composition and stroke origin in mechanical thrombectomy patients: analysis of the stroke thromboembolism registry of imaging and pathology. *Journal of neurointerventional surgery*, 13(7):594–598, 2021.

[9] Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

[10] Bruce CV Campbell, Deidre A De Silva, Malcolm R Macleod, Shelagh B Coutts, Lee H Schwamm, Stephen M Davis, and Geoffrey A Donnan. Ischaemic stroke. *Nature reviews Disease primers*, 5(1):70, 2019.

[11] Marc-André Carbonneau, Veronika Cheplygina, Eric Granger, and Ghyslain Gagnon. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*, 77:329–353, 2018.

[12] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

[13] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[14] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020.

[15] Olivier Dehaene, Axel Camara, Olivier Moindrot, Axel de Lavergne, and Pierre Courtiol. Self-supervision closes the gap between weak and strong supervision in histology. *arXiv preprint arXiv:2012.03583*, 2020.

[16] Victor J Del Brutto, Seemant Chaturvedi, Hans-Christoph Diener, Jose G Romano, and Ralph L Sacco. Antithrombotic therapy to prevent recurrent strokes in ischemic cerebrovascular disease: Jacc scientific expert panel. *Journal of the American College of Cardiology*, 74(6):786–803, 2019.

[17] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[18] Valery L Feigin, Benjamin A Stark, Catherine Owens Johnson, Gregory A Roth, Catherine Bisignano, Gdiom Gebreheat Abady, Mitra Abbasifard, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Vida Abedi, et al. Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet Neurology*, 20(10):795–820, 2021.

[19] James Foulds and Eibe Frank. A review of multi-instance learning assumptions. *The knowledge engineering review*, 25(1):1–25, 2010.

[20] Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, et al. Pre-trained models: Past, present and future. *AI Open*, 2:225–250, 2021.

[21] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[22] Joanna Huang, Murray C Killingsworth, and Sonu MM Bhaskar. Is composition of brain clot retrieved by mechanical thrombectomy associated with stroke aetiology and clinical outcomes in acute ischemic stroke?—a systematic review and meta-analysis. *Neurology International*, 14(4):748–770, 2022.

[23] Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In *International conference on machine learning*, pages 2127–2136. PMLR, 2018.

[24] Umair Akhtar Hasan Khan, Carolin Stürenberg, Oguzhan Gencoglu, Kevin Sandeman, Timo Heikkinen, Antti Rannikko, and Tuomas Mirtti. Improving prostate cancer detection with breast histopathology images. In *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*, pages 91–99. Springer, 2019.

[25] Tae Jung Kim, Ji Sung Lee, Jae Sun Yoon, Mi Sun Oh, Ji-Woo Kim, Soo-Hyun Park, Keun-Hwa Jung, Hyun Young

Kim, Jee-Hyun Kwon, Hye-Yeon Choi, et al. Optimal use of antithrombotic agents in ischemic stroke with atrial fibrillation and large artery atherosclerosis. *International Journal of Stroke*, page 17474930231158211, 2023.

[26] Young-Gon Kim, Sungchul Kim, Cristina Eunbee Cho, In Hye Song, Hee Jin Lee, Soomin Ahn, So Yeon Park, Gyungyub Gong, and Namkug Kim. Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections. *Scientific reports*, 10(1):21899, 2020.

[27] Dawn O Kleindorfer, Amytis Towfighi, Seemant Chaturvedi, Kevin M Cockroft, Jose Gutierrez, Debbie Lombardi-Hill, Hooman Kamel, Walter N Kernan, Steven J Kittner, Enrique C Leira, et al. 2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack: a guideline from the american heart association/american stroke association. *Stroke*, 52(7):e364–e467, 2021.

[28] Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2021.

[29] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[30] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[31] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature biomedical engineering*, 5(6):555–570, 2021.

[32] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[33] Tatsat R Patel, Briana A Santo, TaJania D Jenkins, Muhammad Waqas, Andre Monteiro, Ammad Baig, Elad I Levy, Jason M Davies, Kenneth V Snyder, Adnan H Siddiqui, et al. Biologically informed clot histomics are predictive of acute ischemic stroke etiology. *Stroke: Vascular and Interventional Neurology*, 3(2):e000536, 2023.

[34] Hans Pinckaers, Bram Van Ginneken, and Geert Litjens. Streaming convolutional neural networks for end-to-end learning with multi-megapixel images. *IEEE transactions on pattern analysis and machine intelligence*, 44(3):1581–1590, 2020.

[35] Liyuan Pu, Li Wang, Ruijie Zhang, Tian Zhao, Yannan Jiang, and Liyuan Han. Projected global trends in ischemic stroke incidence, deaths and disability-adjusted life years from 2020 to 2030. *Stroke*, 54(5):1330–1339, 2023.

[36] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001.

[37] J David Spence. Cardioembolic stroke: everything has changed. *Stroke and vascular neurology*, 3(2), 2018.

[38] Senna Staessens and Simon F De Meyer. Thrombus heterogeneity in ischemic stroke. *Platelets*, 32(3):331–339, 2021.

[39] ME Wolf, T Sauer, A Alonso, and MG Hennerici. Comparison of the new asco classification with the toast classification in a population with acute ischemic stroke. *Journal of neurology*, 259(7):1284–1289, 2012.

[40] Weijian Xu, Yifan Xu, Tyler Chang, and Zhuowen Tu. Co-scale conv-attentional image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9981–9990, 2021.