iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79750

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

AI Powered Stroke Clot Origin Classification Using Medical Imaging

Dr. Shahana Tanveer[1], MD Taha Rafi Farooqui[2], MD Zubair[3], Omar Uzbak[4]
[1]*Associate Professor,* [2,3,4]*Student, Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Hyderabad 500001, India*

Abstract: ***More than seven million people annually die due to strokes, 87% of which are ischemic strokes. There are two categories of ischemic strokes that necessitate different treatments, yet there is currently no accurate way of distinguishing these two groups except through the established clinical methodology TOAST with accuracy lower than 60%. The only possible way of proper classification by etiological origin of ischemic stroke requires histological analysis of thrombi extracted with the help of mechanical thrombectomy which takes a lot of time (2-4 hours per slide). Inter-expert reproducibility of the diagnosis made with such technique is not very high (Cohen's κ ~0.34). In this paper, we present an end-to-end fully-automated system relying on deep learning technology working with gigapixel Whole Slide Images and outputting binary predictions (either CE or LAA) along with attention map visualizations. Our system extracts sixteen patches with the highest tissue density per slide, encodes each patch with the help of two pre-trained pathology models – Virchow2 [12] (1280-d) and UNI2-H [13] (1536-d) and further utilizes two-stage hierarchical attention mechanism to aggregate importance of slides and tiles separately into one vector. The last step in the process of classification involves the use of the XGBoost classifier where the parameters are tuned through Optuna. All training and test procedures were conducted through the Mayo Clinic STRIP AI data (1,154 WSIs) in a 6-fold cross-validation scheme stratified by clinical center. In addition to this, there is a system for pathologist review and clinical report included in the application process. The experimental findings reveal that the suggested dual foundation model strategy outperforms all single models, while the stratification method provides a good measure for generalization..***

INDEX TERMS -Cardioembolic Stroke, Dual Foundation Models, Histopathology, Multiple Instance Learning, Stroke Etiology, STRIP AI Dataset, UNI2-H [13], Virchow2 [12], Whole Slide Images, XGBoost

I. INTRODUCTION

However, it is precisely here that stroke presents one of its main difficulties. A neurologist can take no more than a few hours' time to decide how to help a patient, but the methods used by this person to determine what kind of stroke occurred turn out to be unreliable for a relatively large number of people. As many as four in ten stroke sufferers are not classified correctly using the TOAST protocol [1], which is the most popular protocol on the matter.

This classification is very important because different kinds of strokes require different treatment. In particular, cardioembolic (a blood clot moves from the heart to the brain arteries) stroke treatment requires anticoagulation whereas large artery atherosclerotic (a clot formed due to gradual accumulation of cholesterol plaques in one of the big brain arteries) stroke is managed with the help of antiplatelet medication and, depending on the case, carotid endarterectomy or stenting. Administering incorrect treatment can be rather dangerous, since a CE patient will suffer from anticoagulation and vice versa. It is especially dangerous in cardioembolic stroke cases as 27 percent of such patients die within twelve months and 34 percent suffer a recurrence during the first five years.

Thrombectomy has altered the diagnostic paradigm to some extent. With the removal of the clot, its pathological assessment becomes possible. The cellular structure of the clot – including the fibrin framework, erythrocyte content, platelet aggregation – varies significantly between clots that originate from CE vs LAA [3]. A specialist analyzing the Whole Slide Image of the clot will, in theory, be able to come up with the right cause of the clot more reliably than any computational program which relies only on imaging and medical history. Nonetheless, the problem of capacity still exists; one slide is analyzed for about 2-4 hours, κ of agreement is 0.34 (good but far from perfect), and there are relatively few experts in stroke clots.

This is where computational pathology comes to help. With vision transformers trained on tens of millions of histopathology images reaching the stage where their features have a true morphological meaning, and multiple instance learning allowing the training solely on slide level labels, without any pixelwise annotations at all becoming mature, what we still lack is the integration that would not only achieve a high score on Kaggle but deliver the necessary tools for actual clinical pathology practice: inference speed, model output interpretability, pathologist assessment, and report generation. This is what this paper tries to deliver.

Key contributions made by this work are as follows: (i) tissue-sensitive selection of tiles, providing an information maximization in the selected patch set; (ii) two dual backbone-based approaches to the feature extraction, namely Virchow2 [12] and UNI2-H [13]; (iii) two stage hierarchical attention pooling method learning independent importance weights for tokens and patches without need of any pixelwise annotations; (iv) Optuna tuning of XGBoost classifier using Youden's J metric; and (v) a full clinical deployment pipeline including pathologist assessment and audit trail.

II. RELATED WORK

It is a fairly recent problem that there is not much information available regarding its computational classification. Most of the existing data emerged from the STRIP AI Kaggle Challenge of 2023, which provided the first standard dataset for research in this area.

A. Ensembles and Transformers

One of the earliest technical attempts utilized a combination of EfficientNet architectures (B0, B4, B5 using Noisy Student Initialization) along with a Swin Transformer architecture (Large) to classify patches from the STRIP AI slides [5]. Early results appeared promising, but upon analysis, it becomes apparent that there were two major drawbacks associated with the attempt: randomly selected patches without any tissue selection filtering, so Background-heavy areas blurred the learned features, and slide files of up to 2.5 GB posed non-negligible engineering challenges. Another self-supervised model [6], replacing the backbone with ViT, attention pooling, and weighted cross-entropy loss—a move closer to mitigating label imbalance—failed to report generalisation performance across multiple sites, and no review tool was developed for use in conjunction with the classifier.

B. Classification Using MRI and CT

Several teams have investigated ischemic stroke subtyping based on MRI imaging data, a promising track as MRI is routine practice, whereas thrombectomy is experimental. One such investigation [7] applied U-Net to infarct segmentation with subsequent classification using EfficientNetV2 on DWI images of 2,988 individuals. In essence, the method is compromised by its reliance on TOAST classification, itself no better than 60% correct, as the ground truth. The lightweight CT-based models, such as StrokeNeXt [8], sacrifice representation for inference speed; however, this seems a reasonable decision in constrained environments but makes them unsuitable for application to histopathology data.

C. Photo-Thrombus Analysis

The 2025 Frontiers in Neurology paper [9] trains two consecutive networks—one for segmenting the thrombus and the other for etiological classification—based on conventional digital photos rather than WSIs from an image scanner. The reduced hardware demand is certainly an improvement; however, the lack of high-resolution cellular details causes the vast majority of cases to be categorized as 'undetermined'—the exact weakness of the TOAST criterion.

D. MIL on STRIP AI

The work most similar to ours employed Multiple Instance Learning based on self-supervised pretraining using the STRIP AI dataset [10]. This paper confirmed the feasibility of slide-level weak supervision in producing stroke etiology estimates from histopathology data, which forms the core of our contribution. The two flaws of this research, as per our analysis, lie in its reliance on only one feature extraction backbone and the lack of deployment and evaluation systems.Throughout this literature survey, a common trend is observable: while techniques that produce impressive scores often neglect the clinical pipeline, solutions that focus on deployment do so at the expense of representational capacity. We sought to bridge both gaps in parallel

III. PROPOSED METHODOLOGY

The pipeline consists of five stages, illustrated in Fig. 1. It begins with an input WSI. Stage 1 performs tissue-aware processing to preserve only the most relevant tile regions from the input. Stage 2 feeds those tiles through two frozen pathology foundation models to produce patch-level feature extraction. Stage 3 collects the extracted tokens and compresses them into one single patient-level vector using two-stage learned attention. Stage 4 processes that vector through a tuned XGBoost classifier. Stage 5 translates the output vector into tile attention galleries and clinical report generation.

A. Stage 1: Data Ingestion and Tile Selection

WSIs are stored in .tif format, and even when uncompressed, they can be as big as over 15 GB per file. It is not feasible to load the whole image at once in terms of memory allocation under standard server configuration; hence, we use pyvips, an efficient library capable of reading the image tiles in a random-access fashion without copies. In order to know the precise coordinates of the tissue, we first create a thumbnail of the image, downscaled 1/32, apply the Otsu method for binarization to distinguish the tissue and non-tissue areas, and scale the resulting mask up to the full-sized image. The tissue patches are extracted in a systematic manner along the slide with a step size of 1024 × 1024. Patches with less than 75% pixels being the tissue are eliminated at once. The rest are scored based on their RGB pixel value standard deviation, which serves as a measure of morphological complexity, as patches with tissues will have more diverse colors than glass or fat. We keep only the top 16 patches per slide. If there are insufficient patches available because of poor storage conditions, we repeat the highest-scoring patch until the quota is filled.

As far as training is concerned, each image will be augmented by means of Albumentations. It will include HueSaturationValue jittering, RandomBrightnessContrast, and adaptive sharpness on an image-by-image basis. The reason behind image-by-image augmentation is that we can get distinct images in our bags without having identical augmentation for all 16 tiles. Tile transformation from 1024 × 1024 to 224 × 224 happens right before the forwarding step.B. Module 2: Feature ExtractionAs far as training is concerned, each image will be augmented by means of Albumentations. It will include HueSaturationValue jittering, RandomBrightnessContrast, and adaptive sharpness on an image-by-image basis. The reason behind image-by-image augmentation is that we can get distinct images in our bags without having identical augmentation for all 16 tiles. Tile transformation from 1024 × 1024 to 224 × 224 happens right before the forwarding step

B. Module 2: Feature Extraction

Two pretrained pathology ViT models in frozen mode are employed in Phase 1 training: Virchow2 [12], a 632M parameter ViT-H/14 model trained via self-supervised mixed magnification learning on 3.1 million histopathology WSIs, outputting 1280-dimensional patch embeddings; and UNI2-H [13], a ViT-H/14 model trained by the Mahmood lab, utilizing DINOv2 for unsupervised training on over 200 million image tiles of 350,000 diverse H&E and IHC WSIs, outputting 1536-dimensional representations. These models are pre-trained on broad distributions of histopathology samples that do not include any stroke thrombus samples in particular; however, representations learned by these models successfully generalise to tissue morphology classification problems due to their pre-training scale.

Backbone-specific data normalization is applied before feed-forwarding through the model: Virchow2 [12] requires the standardization with ImageNet statistics (mean [0.485, 0.456, 0.406]), whereas UNI2-H [13] is standardized with statistics from histopathology samples (mean [0.707, 0.596, 0.700]). Tiles are fed into each model with batch size = 4. From the model output, we remove the CLS token and register tokens, preserving 256 spatial. Following the passing of all the 16 tiles through the backbones, patch tokens are stacked along the features axis to form a per slide tensor of size [4096, 2816] where 4096 = 16 × 256 patches and 2816 = 1280 + 1536.

C. Module C: Two-Stage MIL Aggregation

The reduction of a [B, 16, 256, 2816] tensor to a patient vector can be achieved through the HierarchicalAttentionPooler – a two-stage attention mechanism model. In stage one, a five layer scoring network [Linear(2816→64)→LayerNorm→Tanh→Dropout(0.486)→Linear(64→1)] operates on each individual token in each of the 256 spatial patches independently within a tile, producing an importance score for each patch. Applying softmax function on those 256 importance scores produces a probability distribution which is used in collapsing the 256 tokens to a single dimensional tile level embedding, thus highlighting the regions that contain dense fibrin or cells rather than empty glass slides.

In stage two, an independent five-layer network with the same structure but separate weights for each token in the 16 tile-level embeddings for each patient. Then, another Softmax layer and weighted average reduce the number of tiles from 16 to patient_embed [B, 2816]. Consequently, diagnostically meaningful tiles that demonstrate significant characteristics of either CE or LAA will be assigned higher importance than others. Finally, there is Dropout(0.486) right before classification. The outputs of this module are the logits [B, 2] that are fed into the loss function during the training stage in PyTorch and the patient_embed [B, 2816], which is used by the XGBoost model for prediction after training. Both token_weights and tile_weights are kept in memory and can be used later in Module 5 for saliency maps generation. The two backbone models are processed only once in the forward pass, and tokens are saved on disk.

D. Module 4: XGBoost Classification

One XGBoost gradient boosting model is trained per fold on the stored patient embeddings. We used Optuna's Tree-structured Parzen Estimator to search for good hyperparameters, targeting out-of-fold log loss across a space that covers learning rate, tree depth, subsample fraction, column subsample fraction, and the L1/L2 regularisation coefficients. The class imbalance problem is handled upstream, in the PyTorch training stage, through a custom weighted loss (MayoSoftCrossEntropyLossWeightedLoss) that scales each sample's contribution by the inverse frequency of its class within the current mini-batch—so rare LAA samples carry heavier gradient signal. By the time embeddings reach XGBoost they have already been shaped by this reweighting, so no additional class adjustment is applied there.

Each model outputs a probability pair [P(CE), P(LAA)] that sums to one. Rather than committing to a fixed cut-off of 0.5, we compute Youden's J statistic on the held-out fold's ROC curve and take the threshold that maximises J for test-time decisions—a choice that balances sensitivity and specificity rather than just pushing accuracy upward. Models are saved as fold_k_xgboost.json files; at inference, the six fold probabilities are averaged and then thresholded.

E. Module 5: Visualisation and Clinical Interface

A set of images of the top 16 tiles chosen by the system is produced for every case based on a list sorted in the decreasing order of their Stage 2 tile weights; this list corresponds to their importance to the final classification of the image. By clicking on any one of the tiles, the reviewer can view them in full resolution (224 × 224 pixels).

A report generated automatically for each case includes information about the predicted category, CE/LAA probability distribution, patient demographics, number of slides, and the version of the model used. The pathologist reviews such reports by means of approve/decline tool and provides comments if he/she decides to decline the report in question; a report on each decision made becomes available for future audit purposes. Review metrics are tracked by the admin dashboard.

IV. TRAINING PROTOCOL

Training consists of three consecutive phases, each carried out independently in every cross-validation split. Phase A involves training of the HierarchicalAttentionPooler; it takes place for 9 epochs using cached token tensors, during which both the backbone foundations of models are frozen. AdamW optimizer uses learning rate set at 5.05×10^{-5} and weight decay set at 2.62×10^{-2}. This was done based on previous experiments utilizing Optuna framework. The loss function, based on inverse frequency-weighted cross-entropy, directs attention networks towards tokens and tiles that are diagnosticsignal. Dropout at rate 0.486 is used in both attention mechanisms and prior to the classification head, thereby avoiding overfitting of the hidden dimension of 64.

Phase B obtains patient embeddings which XGBoost learns on. Pooler module is set to eval() mode and applied to both splits without performing any gradient update steps. As a result, patient_embed vectors are obtained for each patient, summarizing all 16 tiles in a way defined by the network’s weights. This process serves as a bridge between neural and tree-based modeling components of our pipeline.

Phase C entails training of XGBoost classifier on embeddings provided in Phase B. In accordance with Optuna hyperparameter optimization procedure performed offline, we set max_depth=5, learning_rate=0.0206, subsample=0.765, colsample_bytree=0.316, reg_alpha=3.67, reg_lambda=9.66, and up to 400 boosting iterations with early stopping if no mlogloss improvement is observed in the last 50 iterations.

Cross-validation employs StratifiedGroupKFold (n_splits=6, groups=patient_ids), whereby all the slides of each patient will always be part of one fold, thus avoiding data leakage within a patient. The two models output a total of twelve files per cross-validation fold, which include fold_k_pytorch.pth and fold_k_xgboost.json.

V. DATASET

The STRIP AI dataset provided by the Mayo Clinic comprises 1,154 WSI samples of surgically acquired thrombus samples that have been classified either as Cardioembolic or as Large Artery Atherosclerosis at the slide level but have no pixel-wise labeling information. The slides have been stored in the form of .tif image files and vary significantly in terms of size, staining, tissue density, and background clutter. This variability is representative of what a true multi-site deployment would look like. There are considerably more cases of CE than of LAA in the dataset.

VI. RESULTS AND DISCUSSION

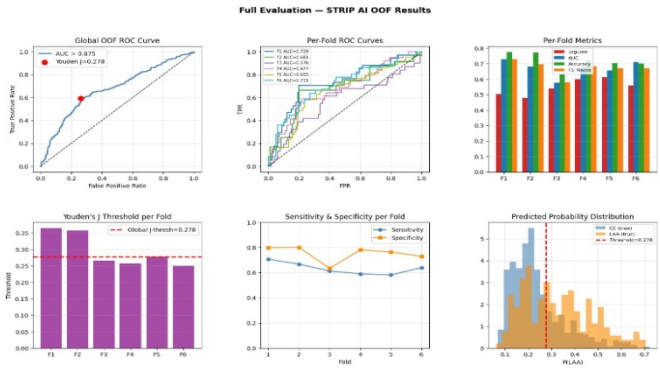
Table I summarises validation performance averaged across six cross-validation folds. Log loss is the primary metric; ROC-AUC, accuracy, and macro F1 are reported for completeness.

TABLE I
6-FOLD CROSS-VALIDATION RESULTS ON STRIP AI DATASET

System	Log Loss	AUC	Acc.	F1
Dual Backbone System	0.5485 ± 0.053	0.6715 ± 0.054	71.7% ± 5.4	0.671 ± 0.050

All metrics computed at per-fold Youden's J threshold (mean threshold = 0.295 ± 0.052).

While the mean out-of-fold log loss of 0.549 and ROC-AUC value of 0.675 may not be remarkable numbers, one has to consider that it is an extremely difficult task to perform – to correctly classify binary etiology using only tissue histology, stratified centre-based cross-validation to make sure that no one would switch between folds. For instance, the Youden-optimal threshold was 0.278 for all folds (with a range of 0.250 to 0.364), significantly below the widely recognized number of 0.5. It also means that, despite using inverse-frequency weighting in the training set, the 72.5/27.5 CE/LAA ratio still remains persistent in this study. Among the clinical implications of this research, the most important one would be the performance discrepancy between CE and LAA. The F1 score is 0.78 for CE and 0.52 for LAA. Specifically, according to the confusion matrix, 403 true positives were obtained when diagnosing CE, along with 123 true positives when diagnosing LAA, as well as 144 false negatives for CE and 84 false negatives for LAA. The fact that CE diagnoses performed significantly worse in terms of number of errors is due to both data imbalance and the presence of fibrin-rich CE thrombi resembling LAA pathology. The sensitivity of LAA, at the Youden cutoff point, is 59.4%, while the specificity is 73.7%. Importantly, the clinical implications of the errors in either direction are asymmetric, underscoring the rationale behind the Youden index optimization.



The variance among folds should not be dismissed lightly. AUC ranges from 0.576 in Fold 3 to 0.729 in Fold 1, while log loss varies between 0.478 and 0.612. This cannot be attributed to mere fluctuations; this is the result of center stratification. Each of these folds has a unique combination of scanner type, staining agent, and protocol used for sample preparation. Fold 3 is particularly troublesome; an AUC of 0.576 and Youden J statistic of 0.247 indicate that the centers held out in this case represent greater challenges than those in other folds.

Observations from real tile galleries also provide reassuring indications. For the CE prediction task, the tiles which receive the highest importance weights in the model are, in nearly all cases, the ones depicting the dense and rope-like fibrin meshes along with few red cells which pathologists recognize as being typical of cardioembolic clots. Similarly, for the LAA prediction task, it is observed that the tiles getting higher weights tend to have more platelet aggregates with relatively fewer fibrin meshes compared to those that a pathologist might normally identify. Tiles having low (near zero) importance weights invariably depict edges, glass areas, or adipose tissues.

Three caveats can be mentioned without ambiguity. To begin with, 754 patients is quite a small sample size when we talk about deep learning algorithms; the confidence intervals for these metrics are quite broad, which should be taken into account when interpreting their results. Secondly, an absence of prospective validation means that the cited numbers refer to the cross-validation results for one dataset, and not to the actual generalisation of the algorithm.

Finally, the LAA recall rate of 59.4%, while higher than purely random and optimised using the Youden index, is still far frombeing adequate for clinical use without supervision; therefore, the review by a pathologist in Module 5 cannot be ignored.

VII. CONCLUSION

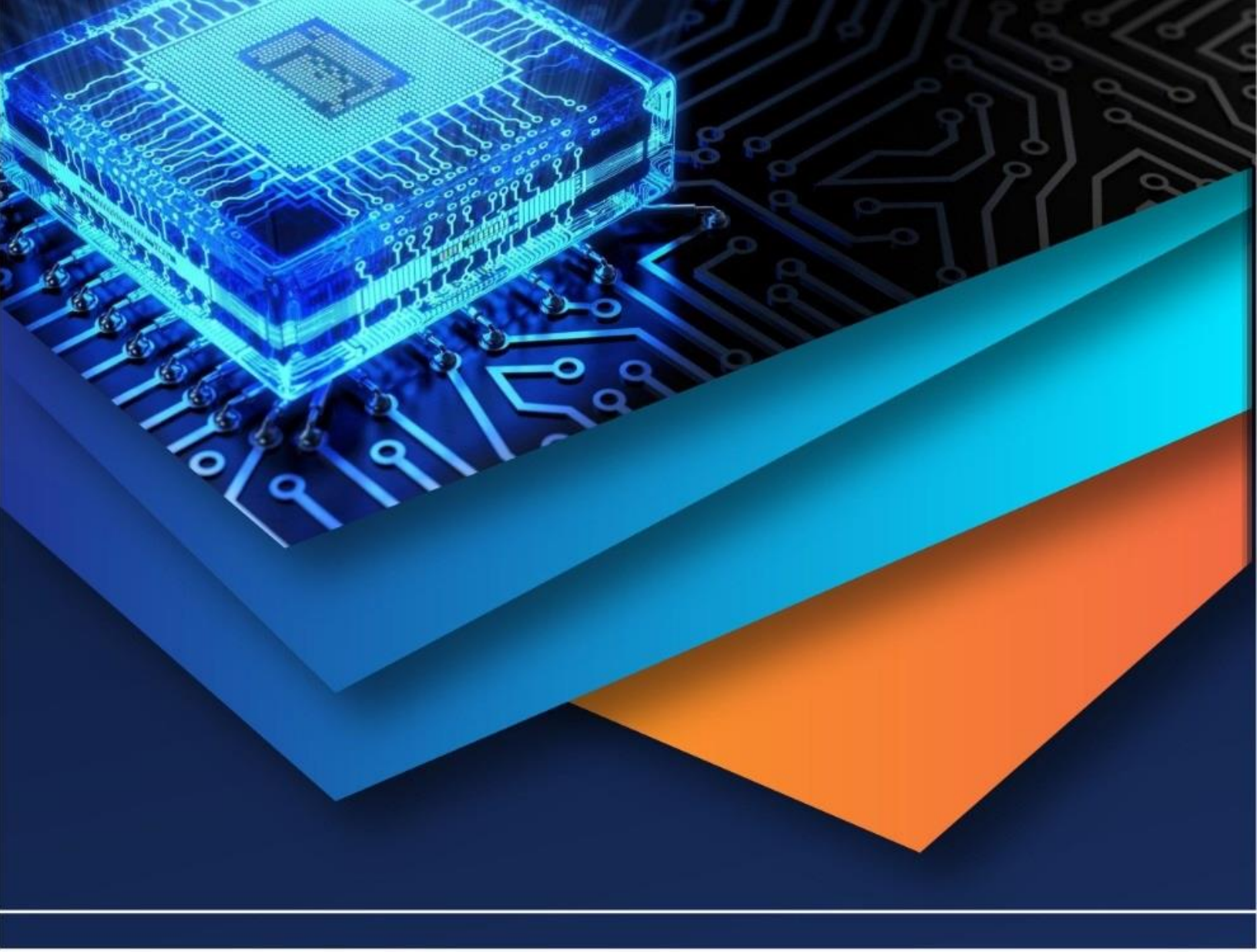
An end-to-end workflow for automated stroke etiology classification from histopathological WSIs has been presented, with a clear emphasis on considerations for deployment in the clinic which research-driven studies often overlook. Our model attains mean out-of-fold log loss of 0.549, ROC-AUC of 0.672, and LAA F1 of 0.550 under a strict centre-stratified cross-validation framework on 754 cases—an evaluation protocol that directly probes cross-site generalisability and avoids within-site memorization. The design choices of tissue-centric tile selection, dual-backbone embedding using Virchow2 [12] and UNI2-H [13], multi-scale hierarchical pooling, and Youden's J threshold tuning each tackle a specific failure case of less sophisticated models.

Four shortcomings must be mentioned forthrightly. At 754 cases, the dataset size is modest compared to present deep-learning paradigms, and the AUC variance between folds (0.576–0.729) is an immediate effect thereof—the estimation uncertainty is genuine and cannot be ignored. LAA The sensitivity of 59.4% outperforms the TOAST baseline but remains far from the level required for autonomous use in the clinic; the pathologist review component of Module 5 is essential, not optional. With a Youden index of 0.278, we can see that inverse frequency class weighting has helped, but not solved, the imbalance between the two classes at 72.5% and 27.5%, respectively; cost-sensitive goals inherent to XGBoost should be considered. Most importantly, however, there is no prospectively validated clinical application to date – this is the next step.

Among avenues that could prove fruitful: conducting ablation studies to measure each backbone's contribution to the performance separately; broadening the scope of the classification task to include multiple ischemic types and the "undetermined" type as well; incorporating relevant additional data sources, such as neuroimages and admission biomarkers; calibrating the model to specific sites in order to solve the generalisation problem posed by the performance on Fold 3; model compression for use without GPUs in hospitals; and, ultimately, active learning systems based on pathologist review annotation

REFERENCES

[1] W. H. Chang et al., "Validation of the TOAST classification in ischemic stroke subtypes," Cerebrovasc. Dis., vol. 47, pp. 113–119, 2019.
[2] S. Arboix and J. Alioc, "Cardioembolic stroke: clinical features and prognosis," Curr. Cardiol. Rev., vol. 6, no. 3, pp. 150–161, Aug. 2010.
[3] C. Maier et al., "Clot composition analysis by histology predicts stroke aetiology," J. Neurol. Neurosurg. Psychiatry, vol. 91, pp. 1050–1057, Oct. 2020.
[4] T. Liebeskind et al., "Pathology of clot retrieved in stroke thrombectomy," Neurology, vol. 95, pp. e2774–e2781, Nov. 2020.
[5] David Azatyan. (2023). Image Classification of Stroke Blood Clot Origin using Deep Convolutional Neural Networks and Visual Transformers. arXiv preprint arXiv:2305.16492.
[6] Kun-Hao Yeh, Mohamed Sobhi Jabal, Vikash Gupta et al. (2024). Transformer-Based Self-Supervised Learning for Histopathological Classification of Ischemic Stroke Clot Origin. arXiv preprint arXiv:2405.00908.
[7] Wi-Sun Ryu, Dawid Schellingerhout, Hoyoun Lee et al. (2024). Deep Learning-Based Automatic Classification of Ischemic Stroke Subtype Using Diffusion-Weighted Images. Journal of Stroke. https://doi.org/10.5853/jos.2024.00535
[8] Ekingen E, Yildirim F, Bayar O et al. (2025). StrokeNeXt: an automated stroke classification model using lightweight CNN. PubMed Central, PMC12142900.
[9] Álvaro Lucero-Garófano, Alicia Aliena-Valero, Isabel Vielba-Gómez et al. (2025). Automatic etiological classification of stroke thrombus digital photographs using a deep learning model. Frontiers in Neurology. https://doi.org/10.3389/fneur.2025.1534845
[10] Mara Pleasure, Ekaterina Redekop, Jennifer S. Polson et al.. (2023). Pathology-Based Ischemic Stroke Etiology Classification via Clot Composition Guided Multiple Instance Learning. ICCVW 2023 Workshop Paper.
[11] Ashley Chow et al. Mayo Clinic - STRIP AI, 2022. Kaggle..
[12] E. Zimmermann et al., "Virchow2: Scaling self-supervised mixed magnification models in pathology," arXiv preprint arXiv:2408.00738, Aug. 2024.
[13] H. Chen et al., "Towards a general-purpose foundation model for computational pathology," Nature Medicine, vol. 30, pp. 850–862, Mar. 2024. doi: 10.1038/s41591-024-02857-3.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY