Classifying Ischemic Stroke Blood Clot Origin: A Comprehensive Study with Ensemble Models

Mohamad Elsayed Eid, Ali Maged, Mohamed Mesilhy, Omar Mohamed Ahmed
Systems and Biomedical Engineering Department, Cairo University

Supervised By:
Inas A. Yassine, PhD
Assoc, Prof.,
Systems and Biomedical Engineering Department, Cairo University
iyassine@eng.cu.edu.eg

Teaching Assistant:
Samar Alaa, MSc
Systems and Biomedical Engineering Department, Cairo University
samar.taher98@eng1.cu.edu.eg

***Abstract*— Accurate classification of blood clot origin in ischemic stroke cases is crucial for effective treatment and prevention strategies. This study presents a comprehensive analysis of deep learning approaches for classifying ischemic stroke blood clots using histopathological images. We implement and evaluate two advanced architectures: PoolFormerV3 and Compact Convolutional Transformers (CCT), comparing their performance in distinguishing between cardioembolic (CE) and large artery atherosclerosis (LAA) origins. Our methodology incorporates extensive data preprocessing, including intelligent patch extraction and augmentation techniques, applied to the Mayo Clinic STRIP AI dataset. The PoolFormer model achieved superior validation performance with 87.52% accuracy and an F1-score of 0.8691, while the CCT model demonstrated computational efficiency with faster training times and achieved 85.53% accuracy. We address challenges related to dataset limitations through robust preprocessing pipelines and transfer learning approaches. Our results demonstrate the potential of ensemble deep learning methods in improving the accuracy and reliability of stroke clot origin classification, contributing to more informed therapeutic decisions in clinical practice.**

I. INTRODUCTION

With about 12.2 million new cases reported each year, stroke is regarded as the world's greatest cause of death and disability. It happens when the blood supply to the brain is cut off, resulting in brain damage and perhaps irreversible neurological disabilities. The most prevalent kind of stroke is an ischemic stroke, which happens when the blood vessel supplying the brain becomes blocked. The cause of this blockage may be a blood clot that has formed in the heart, carotid artery, or other arteries around the body.
Effective stroke prevention and therapy depend on determining the blood clot origin. Recently, there has been a growing interest in creating classification systems for stroke blood clots.

Because of the complexity of the underlying disease, determining the etiology of blood clots that result in strokes is extremely difficult. Numerous conditions, including coagulopathy, atrial fibrillation, and atherosclerosis, can result in blood clots. Because these conditions can interact in complex ways, it can be difficult to determine what caused the clot. The classification method is further complicated by the fact that stroke symptoms might vary depending on the size and location of the clot.

Researchers have been able to develop effective methods for classifying the origin of blood clots that result in strokes thanks to creative ideas. They accomplish this by using a variety of imaging techniques, including MRI and CT, which aid in determining the location and properties of the clot. Additionally, comprehensive clinical examinations that include a full medical history and physical examination, together with laboratory investigations involving blood tests and clotting factor assays, can provide important insights into the underlying reasons that cause clot formation.

II. LITERATURE REVIEW

Numerous studies have explored the application of deep learning, particularly convolutional neural networks (CNNs), for classifying the origins of ischemic stroke blood clots using histopathological images [1], [2], [3], [4], [5].

These models demonstrate exceptional capability in extracting intricate features from high-dimensional image datasets [6], facilitating the differentiation between clot origins such as cardioembolic (CE) and large artery atherosclerosis (LLA) [1], [2], [3], [4]. For instance, Rao et al. [1] introduced an ensemble model integrating EfficientNet-B0, VGG19, and ResNet-152, which outperformed individual models by achieving lower loss values, underscoring the efficacy of ensemble strategies [1].

Similarly, Krishnan et al. [3] adopted a two-stage framework, utilizing MobileNetV3 for segmentation and fine-tuning pre-trained models for classification. Among these, PoolFormer demonstrated superior performance metrics [3]. Nevertheless, the dependency on extensive, high-quality datasets poses a significant challenge [2].

Yeh et al. [2], while showcasing promising results with a transformer-based self-supervised learning method, highlighted the importance of advanced preprocessing techniques to enhance both model performance and interpretability [2]. Moreover, further research is required to improve the generalizability of these models across diverse imaging modalities and datasets [5].

Model Architectures and Ensemble Methods

The selection of deep learning architectures plays a crucial role in determining model performance. Research has investigated various CNN architectures, such as VGG16 [7], ResNet-152 [1], EfficientNet [1], [5], [8], and MobileNetV3 [3], demonstrating the adaptability of deep learning in this field. Additionally, ensemble methods, which combine multiple models to enhance prediction accuracy and robustness, have yielded promising results [1], [8], [9], [10].

For example, Azatyan [9], [10] utilized an ensemble comprising four deep neural network architectures to improve classification accuracy in distinguishing cardiac and large artery atherosclerosis as ischemic stroke origins. This success underscores the potential of ensemble strategies to leverage the strengths of diverse models and mitigate their limitations. However, the computational demands associated with training and deploying ensemble models must be carefully addressed.

Data Augmentation and Preprocessing Techniques

The effectiveness of deep learning models is highly dependent on the quality and volume of the training data. To enhance model robustness and generalization, data augmentation techniques are frequently utilized to expand the training dataset by applying various transformations to existing images [6]. For instance, Hambali and Agwu [6] employed data augmentation along with pixel brightness transformations for image enhancement and adversarial training, achieving up to 97% accuracy through transfer learning [6]. Furthermore, preprocessing methods such as noise reduction and image normalization are critical for optimizing model performance [6]. However, the choice of preprocessing steps often requires customization based on the specific characteristics of the dataset and the architecture of the model [2].

III. Methodology

Research Design

This study adopts a deep learning-based approach for classifying the origins of ischemic stroke blood clots using histopathological images. The methodology focuses on the development and evaluation of both convolutional neural network (CNN) and Vision Transformer (ViT) models to extract complex features from high-dimensional image data. To enhance model robustness and generalization, ensemble learning techniques are employed, combining the strengths of multiple architectures.

The study also investigates preprocessing and data augmentation strategies to address challenges posed by limited datasets and variations in image quality. A standardized evaluation framework is used to assess model performance, ensuring reproducibility and comparability.

Data Collection

Histopathological image data used in this study were sourced from publicly available and institutional datasets, ensuring a diverse representation of ischemic stroke blood clots. The datasets include annotated images detailing clot origins such as Cardioembolic (CE) and large artery atherosclerosis (LLA). To mitigate challenges associated with limited data, data augmentation techniques, including geometric transformations, brightness adjustments, and noise injection, were applied to expand the dataset size artificially. Images underwent preprocessing steps such as normalization to optimize input quality and model performance

IV. Dataset And Features

1) STRIP AI background clot dataset [11]: To create a background classifier that removes empty patches from digital pathology images, we used the STRIP AI background clot dataset from Kaggle. The dataset consists of 19,998 photos, half of which include cell contents and the other half of which are empty patches with no cell content.

2) The Mayo Clinic - STRIP AI dataset: available publicly through Kaggle, is widely utilized in studies focused on classifying the origins of ischemic stroke blood clots [1], [2]. Its popularity is attributed to the large and readily accessible data, it offers for training and validating deep learning models.

Dataset Limitations

However, detailed and standardized descriptions of the dataset's composition are often missing in the literature.

Variability in reporting the dataset's specifics, such as the total number of whole slide images (WSIs), types of clots represented (e.g., cardioembolic, large artery atherosclerosis, and other etiologies), and imaging modalities, limits its comprehensive evaluation. This inconsistency poses challenges in determining the dataset's suitability for different deep-learning models and training methodologies [2].

Moreover, the potential biases inherent in the dataset remain underexplored, raising concerns about the generalizability of models trained on this data. For instance, demographic information such as patient age, sex, and comorbidities is not thoroughly characterized. Addressing these gaps through detailed reporting and dataset characterization could significantly enhance its utility and reliability for the research community.

Data Pre-Processing

Data preprocessing is a critical step in preparing high-quality datasets for analysis. This section outlines the preprocessing steps applied to the STRIP AI clot dataset to ensure its suitability for training and evaluation.

1. Image Extraction

The STRIP AI dataset comprises whole-slide images (WSIs) of digital pathology images in TIFF format, a commonly used file type in the field. Due to the substantial size of these images (total of around 200GB), we first resized them by a factor of 1/10 in both height and width to reduce computational complexity while retaining essential structural details. We then employed the OpenSlide package to efficiently extract smaller patches of $600 \times 600 \times 3$ pixels

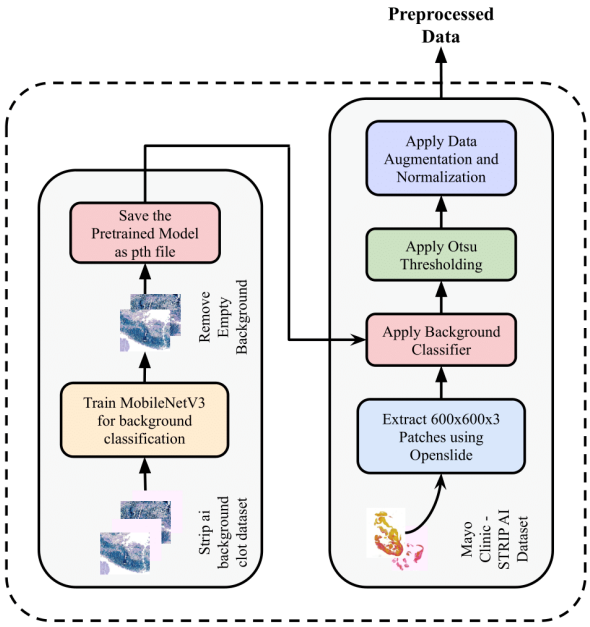

Figure 1 Data preprocessing Pipeline

from the resized images. These patches were subsequently saved in PNG format to facilitate downstream analysis.

2. Removal of Irrelevant Patches

We first trained a MobileNetV3 model on the STRIP AI background clot dataset [11], then used this pre-trained model classifier to exclude white or irrelevant images from the dataset. This step ensured the dataset contained high-quality patches relevant to the classification of ischemic stroke blood clots.

Subsequently, we applied Otsu's thresholding technique to filter out low-quality patches. This method involved calculating the area of the slide contents relative to the total image area and retaining patches only if the slide content occupied more than 30% of the total area. This step effectively removed patches with insufficient or irrelevant content, ensuring that only meaningful data was retained for further analysis.

After applying both the background classifier and Otsu's thresholding, we successfully eliminated empty and low-quality patches. Consequently, the data size was reduced from 124,800 patches to 118,600 patches.

3. Data Splitting and Augmentation

To avoid data leakage and maintain the integrity of the evaluation process, the dataset was first split into 80% training data and 20% testing and validation data. Augmentation was then applied exclusively to the training dataset using a carefully designed augmentation pipeline implemented with the Albumentations library. The pipeline included the following transformations:

- All images were resized to 224 × 224 pixels.
- Horizontal and vertical flipping, random 90-degree rotations, and brightness and contrast adjustments were applied, each with a 50% probability.
- Shift-scale-rotate transformations were applied with shift limits of 5%, scale limits of 5%, and rotation limits of ±15 degrees.
- Normalization was performed using the standard ImageNet mean ([0.485, 0.456, 0.406]) and standard deviation ([0.229, 0.224, 0.225]).

For the testing and validation datasets, a simpler pipeline was employed, consisting only of resizing to 224 × 224 pixels and normalization using the same ImageNet statistics. All images were then converted into PyTorch tensors using. These transformations ensured the training data was sufficiently augmented for robust learning while maintaining consistent dimensions and distributions across all datasets.

V. MACHINE LEARNING PIPELINE

MODEL SELECTION

After preparing the data loaders for training, testing, and validation, we proceeded to build the classification model. For this task, we selected

1) PoolFormerV3 architecture: based on its promising results as highlighted in the literature. Given the constraints of our dataset size and computational resources, training the model from scratch was not feasible. Thus, a transfer learning approach was employed.

2) Compact Convolutional Transformers (CCT): We chose this model for its ability to effectively handle small datasets through its hybrid architecture, which combines convolutional layers for local feature extraction with transformer layers for global context modeling. CCT's parameter efficiency and robust generalization make it well-suited for medical imaging tasks where data is often limited, aligning perfectly with the Mayo Clinic dataset's characteristics.

EVALUATION CRITERIA

To optimize this problem, we apply the Weighted Multi-Class Logarithmic Loss (WMCLL) loss function.

$$\text{WMCLL} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} w_c \cdot y_{i,c} \cdot \log(\widehat{y_{i,c}})$$

Where:

$-$ (N): $Total\ number\ of\ samples.$

$-$ (C): $Number\ of\ classes.$

$-$ (w_c): $Weight\ assigned\ to\ class(c).$

$-$ ($y_{i,c}$): $Binary\ indicator\ (0\ or\ 1)\ if\ class\ (c)\ is\ the\ correct\ class\ for\ sample\ (i).$

- ($\widehat{y_{i,c}}$): $Predicted\ probability\ of\ sample\ (i)\ belonging\ to\ class\ (c).$

Table 1 COMPARATIVE PERFORMANCE ANALYSIS OF STROKE BLOOD CLOT ORIGIN CLASSIFICATION MODELS

Model	Weighted Multi-Class Logarithmic Loss	Accuracy	Precision	Recall	F1-Score
Results from" Image Classification of Ischemic Stroke Blood Clot Origin"					
Best Stacked Model (EfficientNet-B0, VGG19, ResNet-152)	0.69312	-	-	-	-
Results from Darapaneni et al. [5]					
ResNet152	-	0.7484	-	-	-
DenseNet121	-	0.6821	-	-	-
EfficientNet	-	**0.7615**	-	-	-
Our Results					
PoolFormer	**0.5576**	**0.8752**	**0.8643**	**0.8752**	**0.8691**
CCT	0.6812	0.8553	0.8513	0.8553	0.8532

This loss function enhances the usual multi-class logarithmic loss with weights for each class. The weights change each class's contribution to the overall loss, highlighting underrepresented or important classes. WMCLL penalizes erroneous predictions based on the set class weights. The loss is estimated using the logarithm of expected probabilities, accounting for confident but inaccurate forecasts.

Classification models are evaluated using metrics like accuracy, precision, recall, and F1 score. Accuracy measures the proportion of correctly identified samples and is suitable when class sizes are comparable. Precision evaluates the percentage of true positives among predicted positives, making it crucial when false positives are costly. Recall assesses the proportion of true positives among actual positives, emphasizing importance when false negatives are critical. The F1 score, a harmonic mean of precision and recall, is particularly useful for imbalanced datasets.

MODEL TRAINING

To adapt the model to our specific classification task, the last fully connected (FC) layer of the original model was removed and replaced with a custom FC layer designed to output two logits corresponding to our task. Dropout and batch normalization layers were also incorporated into the architecture to reduce the risk of overfitting. During training, the earlier layers of the model were frozen to retain the learned representations from the pre-trained weights, while the last layers were fine-tuned to better capture the task-specific features.

During training, the model was optimized using the AdamW optimizer with weight decay and a dynamic learning rate scheduler. The training was divided into epochs, each of which had a training and validation phase. During the training phase, the model analyzed mini-batches of data to update weights depending on the loss, while the validation phase assessed the model's performance using measures such as accuracy, precision, recall, and F1-score, which were calculated using macro averaging to account for class imbalance.

To prevent overfitting, early stopping was employed based on the validation loss or F1-score, with checkpoints saving the best-performing model. Metrics and loss trends were logged for analysis, and the confusion matrix was generated to evaluate classification performance.

VI. IV. RESULTS

Table I shows the competition for stroke blood clot origin categorization using different deep-learning models. Rao et al. [1] used a stacked model of EfficientNet-B0, VGG19, and ResNet-152 to establish a baseline with a Logarithmic Loss of 0.69312, but did not disclose its accuracy or other metrics.

Darapaneni et al. [5] found that the ResNet152 model obtained 74.84% accuracy, whereas regular CNN and EfficientNet matched for the greatest accuracy at 76.15%. DenseNet121 had lower accuracy (68.21%). Although ResNet152 excels in this application, the traditional CNN and EfficientNet models also perform well.

The PoolFormer model demonstrated consistent improvements over training, achieving a validation accuracy of 85.23% and an F1-score of 0.8385 at epoch 10. By epoch 11, the model further improved, with a validation accuracy of 87.52% and an F1-score of 0.8691. These results indicate that the model effectively captured the underlying patterns, as reflected by the balanced precision and recall scores during the training and validation phases.

The CCT model exhibited rapid convergence, with early stopping at epoch 31. At this point, the model achieved a validation accuracy of 85.53% and an F1-score of 0.8532, with marginally better precision and recall than PoolFormer. While the CCT model's training loss was slightly lower than

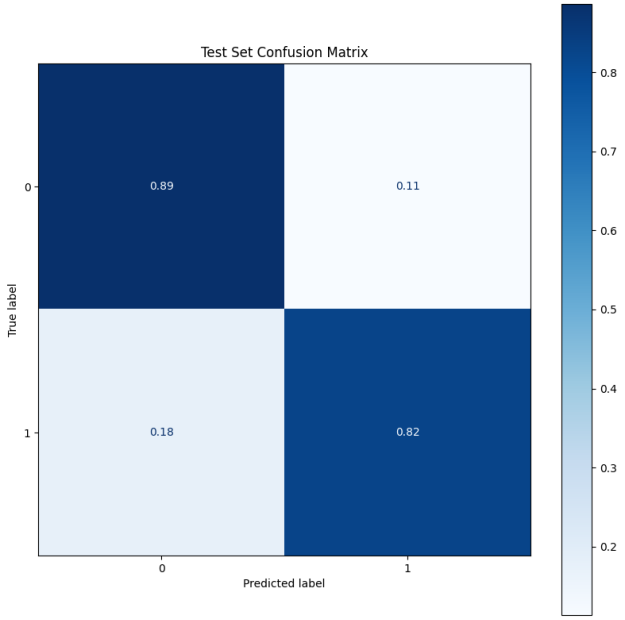

Figure 2 Normalized Confusion Matrix on Test Set for CCT Model

that of PoolFormer, its validation loss showed a slight increase, suggesting potential overfitting.

The PoolFormer model outperformed the CCT model in validation metrics, particularly in F1-score and accuracy. However, the CCT model trained significantly faster, with an average epoch time of ~475 seconds, compared to PoolFormer's ~984 seconds, highlighting its computational efficiency.

The test set confusion matrix (Figure 2) reveals that the model achieved an overall balanced classification performance. For class 0 "CE" the model correctly predicted 89% of the samples, while 82% of class 1 "LAA" samples were correctly classified. The false positive rate for class "0" was 11%, and the false negative rate for class "1" was 18%, indicating that the model slightly underperformed in identifying some samples from class "1."

In summary, PoolFormer provided better validation performance, while CCT excelled in training speed, offering a trade-off between accuracy and computational cost.

V. Conclusion

This study demonstrates the effectiveness of modern deep learning architectures in classifying ischemic stroke blood clot origins using histopathological images. The comparison between PoolFormerV3 and CCT models reveals important trade-offs between accuracy and computational efficiency. The PoolFormer model's superior validation metrics (87.52% accuracy, 0.8691 F1-score) suggest its potential for clinical applications where precision is paramount, while the CCT model's faster training time (475 seconds per epoch versus 984 seconds) offers advantages for rapid deployment and iteration.

The implementation of comprehensive preprocessing techniques, including automated background removal and quality assessment, proved crucial in optimizing model performance. Our approach to handling dataset limitations through careful augmentation and validation strategies addresses common challenges in medical image analysis.

The balanced performance across both cardioembolic and large artery atherosclerosis classifications, as evidenced by the confusion matrix results (89% accuracy for CE and 82% for LAA), indicates robust generalization capabilities. However, the slightly lower performance in LAA classification suggests areas for future improvement, possibly through more targeted data collection or specialized model architectures.

These findings contribute to the growing body of evidence supporting the use of artificial intelligence in stroke diagnosis and treatment planning while highlighting the importance of balanced consideration between model performance and practical implementation constraints in clinical settings.

VI. Future Work

Several key directions for future research could enhance the effectiveness of blood clot classification:

1. Pre-training on Histopathological Images: Instead of using ImageNet pre-trained weights, models could be pre-trained on large-scale histopathological image datasets. This domain-specific pre-training would likely improve feature extraction capabilities and overall classification performance.
2. Multi-modal Integration: Combining histopathological images with clinical metadata and radiological imaging data could provide more comprehensive information for classification.
3. Dataset Enhancement: Expanding the dataset to include more diverse cases from multiple medical centers would improve model generalization and robustness.
4. Clinical Validation: Conducting prospective clinical trials to validate model performance in real-world settings and developing user-friendly interfaces for clinical deployment.
5. Extended Classification: Developing capabilities to identify additional clot origins beyond CE and LAA, including mixed-origin clots.

These improvements, particularly the implementation of histopathology-specific pre-training, could significantly advance automated stroke clot classification and improve patient outcomes through more precise diagnosis.

REFERENCES

[1] Masabathula V S Raghavendra Rao, Surya Puligundla, Nikhil Sai Ekkala, Srilatha Chebrolu. "Image Classification of Ischemic Stroke Blood Clot Origin using Stacked EfficientNet-B0, VGG19 and ResNet-152."

[2] MS Kun-Hao Yeh, MD Mohamed Sobhi Jabal, PhD Vikash Gupta, MD David F. Kallmes, MD Waleed Brinjikji, Barbaros, Dds PhD Selnur Erdal. "Transformer-Based Self-Supervised Learning for Histopathological Classification of Ischemic Stroke Clot Origin."

[3] Koushik Sivarama Krishnan, P. J. J. Nikesh, Swathi Gnanasekar, Karthik Sivarama Krishnan. "Advancing Ischemic Stroke Diagnosis: A Novel Two-Stage Approach for Blood Clot Origin Identification."

[4] Kozhamseitova, Aisha. "Development of the method of blood clot origin classification in ischemic stroke."

[5] Narayana Darapaneni, B. Sudha, A. Reddy, Ab Abdul Karim, Dhanalakshmi Marothu, S.Kulkarni, Deepak Das Menon. "Image Classification of Stroke Blood Clot Origin."

[6] M. A. Hambali, Paul A. Agwu. "Adversarial Convolutional Neural Network for Predicting Blood Clot Ischemic Stroke."

[7] Muhammad Usama Tanveer, Kashif Munir, Bharati Rathore, Abdulatif Alabdulatif, R. Jhaveri, MahamFatima. "Neuro-VGNB: Transfer Learning-Based Approach for Detecting Brain Stroke."

[8] Ritesh Maurya, Rudra Shaurya, T. Gopalakrishnan, Vishnu Srinivasa Murthy Y, Priya S. "Acute Ischemic Brain Stroke Classification Using Attention-Augmented Ensemble Approach."

[9] Azatyan, David. "Image Classification of Stroke Blood Clot Origin using Deep Convolutional Neural Networks and Visual Transformers."

[10] "Image Classification of Stroke Blood Clot Origin Using Deep Convolutional Neural Networks and Visual Transformers: A Kaggle Competition.

[11] alejopaullier. strip-ai-background-clot. 2022. URL: https://www.kaggle.com/datasets/alejopaullier/strip-ai-background-clot