Leveraging Global Context with RPNs and Vision Transformer for Enhanced Cerebral Microbleed Detection in Single-Plane MRI Imaging

A Thesis Proposal

Presented to the

Department of Computer Science

College of Information and Computing Sciences

University of Santo Tomas

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science in Computer Science

By:

Araza, Alessandro Andrei, A.

Entrata, Joshua Kyle, K.

Lorenzo, Zedrick, M.

Sebastian, Nigel Haim, N.

Adviser:

Mr. Jessie James P. Suarez

April 2024

Table of Contents

List of Figures

List of Tables

Chapter I The Problem and Its Background

A. Introduction

Cerebral microbleeds (CMBs) are a chronic collection of microscopic brain hemorrhages within the brain's tissue (Martinez-Ramirez et al., 2014). According to Cordonnier et al. (2007), approximately 5% of the healthy population has microbleeds, but their increased occurrence may be linked to various medical conditions. There has been a growing focus on the connections between CMBs and cognitive dysfunction (Yakushiji et al., 2008). However, CMBs are also observed in elderly individuals without known health issues (Akoudad et al., 2015). These lesions are common in various diseases, such as cerebral amyloid angiopathy, cerebrovascular disease, traumatic brain damage, and Alzheimer's disease (Haller et al., 2018). Compared with the global hemispheres, microbleeds typically occur in deep structures, such as the thalamus, brain stem, basal ganglia, and cerebellum (Viswanathan et al., 2006). In cases with individuals who encountered head trauma, the lesions are usually seen in the frontal, temporal, and parietal lobes (Yates et al., 2014). Therefore, to ensure precise diagnosis and effective treatment, it is very important to accurately distinguish CMBs from structures that appear similar on neuro-MRI, such as iron deposits, calcifications, and veins.

According to the study of Martinez-Ramirez et al. (2014), the paramagnetic properties of CMBs make it possible to detect CMBs in MRI scans. In particular, T2*GRE and SWI MRI are used to detect CMBs because these MRI scans are sensitive enough to detect the rupture of blood vessels as small as 200 micrometers in diameter. They are effective because they provide high contrast between brain parenchyma and highly paramagnetic material, such as deoxyhemoglobin, superparamagnetic, hemosiderin, and diamagnetic calcium (Yates et al., 2014). Additionally, according to Poels et al. (2011), CMBs are easily depicted in T2* Gradient-Recalled Echo (GRE)

or Susceptibility Weighted (SWI) MRI sequences and appear as small circular or elliptical lesions with a size of 2 to 10 mm. However, according to Haller et al. (2018), spotting CMBs can be challenging because of their similar-looking features and artifacts. They also mentioned that in diagnosing CMBs, these "mimics" are incorrectly identified in 11% to 24% of cases. According to them, these other conditions that are mistakenly identified as CMBs include tiny vessel tears (microdissections), small bulges in blood vessel walls (microaneurysms), and small calcium deposits (microcalcifications). In the study of Charidimou et al. (2011), they found that examples of these mimics include blood vessels lying in the subarachnoid space, calcifications of the basal ganglia, and cavernous malformations, all of which can appear as small, dot-like areas of low signal on T2*-GRE MRI scans. They also mentioned that a careful examination of adjacent slices and the use of various imaging modalities (CT, T2 MRI, FLAIR, or DWI), along with assessing the shape and location of the lesions, helps in accurately identifying CMBs. Because of these mimics, the manual detection of CMBs produces a high rate of false positives and is tedious and time-consuming, and the outcomes can vary among neuroradiologists, making the process subjective (Fang et al., 2023).

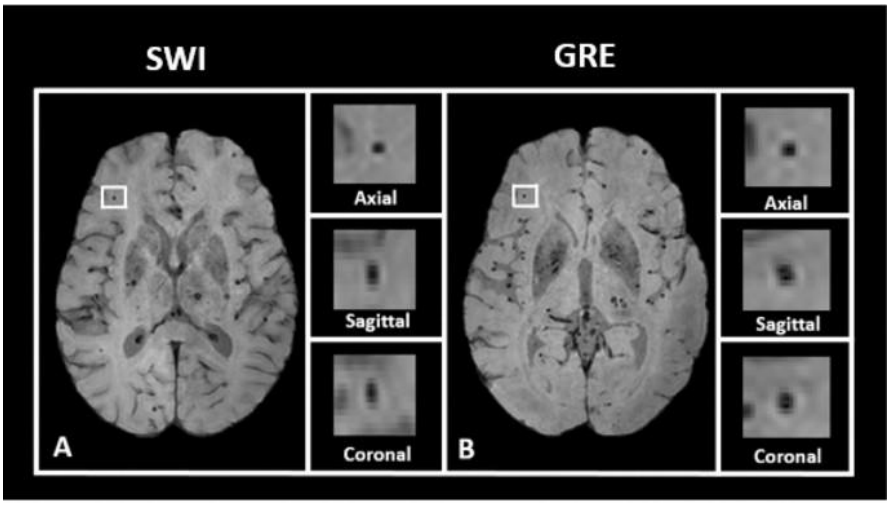


Figure 1.1 Sample microbleed location in different planes

The development of automatic CMB detection is crucial. It is believed that it can alleviate the workload of radiologists and clinicians and enhance the efficiency and reliability of radiation assessments. Present works in CMB detection have been influenced by the integration of machine learning and advanced image processing techniques. Barnes et al. (2011) proposed a semi-automated detection method for CMB. They utilized a threshold algorithm to obtain hypointensities in brain MRI and used a support vector machine (SVM) to separate true CMB from other marked hypointensities within the image. Bien et al. (2013) employed a computer-aided method based on 2D fast radial symmetry transform (RST) to generate potential CMB regions. Afterward, a 3D region growing method was performed on the potential CMB candidates to eliminate false positive CMBs. Chen et al. (2018) employed a novel 3D deep residual network architecture to reduce the false positives that remain after initial detection and improve specificity. Hong et al. (2019) used ResNet-50 to extract features and introduced transfer learning because of their small CMB sample size. Al-masni et al. (2020) utilized YOLO-v2 to identify potential CMB candidates in the first stage and 3D-CNN for the false positive reduction in the second stage. Sundaresan et al. (2023) implemented U-Net for the first stage and 3D-CNN for the second stage, which utilizes the knowledge distillation method.

Previous approaches primarily focused on classifying CMB candidates from a fixed number of adjacent slices, voxel subspaces, or multiple planes. They often fail to fully utilize the surrounding information, which could lead to insufficient detection capabilities and high rates of false positives. This study proposes a novel approach that integrates global brain context within a single plane (axial plane) to address these limitations. By employing attention mechanisms, this study seeks to enhance the accuracy of CMB detection without compromising the precision and

the need for multiple-plane analysis, thereby simplifying the diagnostic process and reducing the risk of missing crucial information.

B. Background of the Study

The detection and analysis of CMBs have become crucial in diagnosing and understanding various neurological diseases. These small spots are critical indicators often associated with increased risks of dementia and cognitive decline in patients with cognitive impairment, thus precise detection is crucial in the field of medicine (Lee et al., 2018). However, Ferlin et al., (2021) mentioned that medical experts often find it challenging to detect CMBs due to their small size and the complex nature of the surrounding brain tissue. They further highlighted that while recent advancements in machine learning have significantly improved medical imaging techniques, challenges still persist.

Several studies have developed models to address these challenges. For example, Chen et al. (2018) developed a 3D Deep Residual network to distinguish true cerebral microbleeds from mimics using fixed-size 3D patches. Their methodology starts with identifying potential CMB candidates through an initial detection algorithm. Although their model achieved a precision rate of 71.9% and significantly reduced false positives by 89%, the fixed-size patches approach may not be optimal for all instances. They could either be too small, potentially losing crucial information for classifying CMBs, or too large, adding irrelevant information that might confuse the model.

Similarly, Fan et al. (2022) developed a model using 3D-CNN modified from a 3D U-Net architecture. The training involves using three consecutive layers of slices to determine the presence of CMBs by analyzing the middle slice for detection. Despite achieving a Dice coefficient of 0.72, a precision rate of 71.8%, and a recall rate of 76.5%, this method limiting the

input to only three consecutive slices might neglect important details if the slices are thick or widely spaced. This limits its applicability to different datasets.

Lee et al. (2022) proposed a Triplanar Ensemble Detection Network (TPE-Det) that leverages 2D Convolutional Neural Networks (CNNs) across axial, sagittal, and coronal planes. This innovative approach improved accuracy and reduced false positives by combining detections across multiple planes, thus utilizing 3D information without the computational cost of 3D CNNs. The model achieved a high sensitivity rate of 96.05%, an average of 0.88 false positives, and a precision rate of 76.76%. It was able to distinguish true positives from false positives by leveraging spatial information from three planes. However, using only the axial plane proved to be insufficient, yielding a high number of false positives with a precision rate of 9.49%, a sensitivity rate of 97.26%, and an average of 9.49 false positives. These results suggest that relying on a single plane diminishes performance compared to using multiple planes.

Kim et al. (2023) introduced a novel 3D deep learning framework that enhances clinical diagnosis support by detecting CMBs and providing their anatomical location. They utilized a U-Net architecture with a Region Proposal Network (RPN) and incorporated Feature Fusion Module (FFM) and Hard Sample Prototype Learning (HSPL) to enhance detection accuracy and reduce false positives. Using the GMC SWI dataset, the model achieved a sensitivity rate of 94.66%, a precision rate of 85.54%, and an average of 0.52 false positives in subjects with CMBs and an average of 0.57 false positives in subjects without CMB.

Most detection models, as described by Ferlin et al. (2021), used a fixed number of adjacent slices to analyze and classify CMBs, which may lead to misclassification and high false positives due to incomplete contextual information of the brain. Applying the anatomical localization task in their model helped them reduce false positives. In the study of Kim et al., it

was observed that even though the detection task classified the candidates as true CMBs based on their shape and position across three consecutive slices, candidates located in regions where CMBs could not anatomically exist were excluded. This task effectively eliminated potential mimics and enhanced the accuracy of the model.

Fan et al. (2022) and Chen et al. (2019) utilized adjacent brain slices and three-dimensional patches centered on CMB candidates, respectively. However, these methods may miss crucial information by focusing on the localized areas only and neglecting broader brain context. Fan et al.'s reliance on fixed count of slices in all samples can be inadequate for accurately identifying microbleeds, as evidenced in the study by Kim et al. (2023). Similarly, Chen et al.'s 3D patches need the same dimensional size for every extraction because they are fed into a neural network where the number of input parameters cannot be inconsistent. This restriction might neglect important information especially when the localized area is too small.

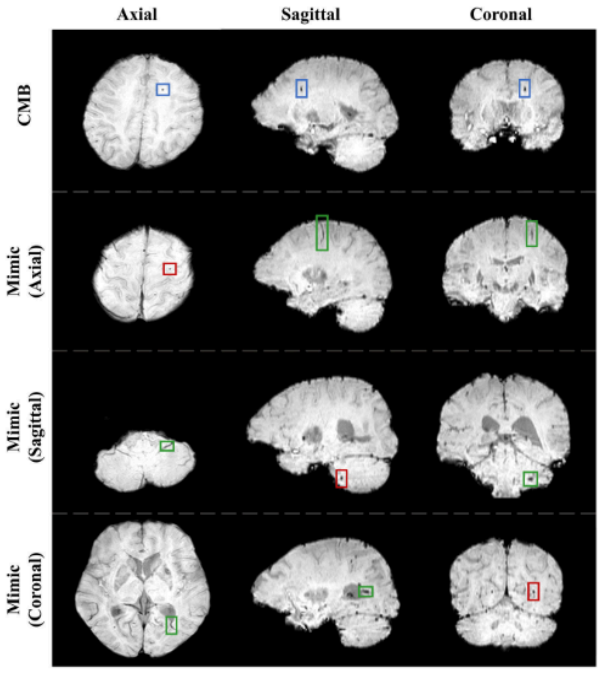


Figure 1.2 Examples of CMB and its mimics on the axial, sagittal, and coronal planes

The challenge lies in developing strategies that leverage the full context of the brain within the same imaging plane to minimize false positives. Lee et al. (2022) observed that CMBs typically appear as spherical shapes in the axial plane and as elongated ellipse shapes in the sagittal and coronal planes. It was also observed that mimics of CMBs have similar intensities and intensities and shapes as CMBs on all three planes. However, these mimics are usually categorized as veins and are only distinguished after examining more slices, as seen in *Figure 1.2*. By integrating more contextual information around the surrounding areas, which are often overlooked by methods using fixed adjacent slices or 3D patches, the introduction of attention mechanisms could significantly enhance the detection of CMBs.

C. Theoretical Framework

The theoretical framework incorporates the outline of different theories and concepts of previous studies to automate microbleed detection. This section provides an overview of studies of different concepts that served as the foundation of our proposed methodology. Figure 1.3.1 illustrates the techniques and methods used in their studies as part of their architecture for detecting microbleeds. It overviews Classification, Localization, Attention Mechanisms, and Preprocessing methods and theories for the study.

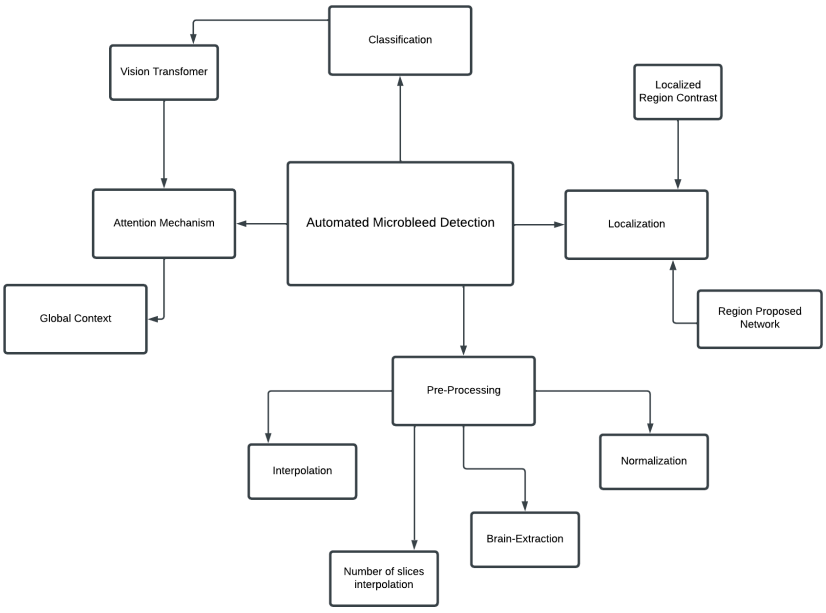


Figure 1.3.1 Theoretical Framework Diagram

A. Localization

Localization detects a specific region of interest (ROI) in an image. Interpretation tasks depend on this stage; therefore, it is a critical prerequisite for all object detection tasks (Alaskar et al., 2022). It is also used in studies of brain tumors, such as the study by Li et al. (2021).

Kim et al. (2023) stated that localization is not only limited to determining CMBs but also reduces false positive results through anatomical information. Another study by Chen et al. (2015) proposed a method where one of the three steps is the localization of CMB candidates with high sensitivity. They analyzed these scans following the production of a binary image through a statistical method. Each candidate's center is the centroid of a 3D growing connected component.

- **Region Proposal Network**

Introduced by Ren et al. (2016) Region Proposal Networks (RPN) shares full-image convolutional features with the detection network, it enables region proposal with free of cost on memory. It is a convolutional network that predicts object bounds and objects scores at each position simultaneously. Their proposal shares convolutional layers with a small marginal cost for computing time.

Kim et al., (2023) integrated RPN and Fast R-CNN into a single network that shares their convolutional features with attention mechanisms. RPN is responsible for instructing the unified network what location to look at. They concluded that RPN is efficient and accurate that shares convolutional features. It improves regional proposal quality and the overall accuracy of their model.

Other studies conducted by Yan et al. (2023) that applied Contrastive learning frameworks that were integrated in the localized region. This is done to enhance self-supervised models for segmentation with limited annotations.

B. Attention Mechanism

Inspired by how a human vision works, the attention mechanism works by dynamically focusing on a subset of the given visual field. This long-time developing mechanism has become a new hope for advanced models. A survey by Xie et al. (2023) stated that the application of attention mechanisms has gained a

steady annual growth from 2017 to 2023. This is because of hierarchical feature learning capabilities. It serves as a plug-in sublayer into the convolution block.

- **Vision Transformer (ViT)**

  Dosovitsky et al. (2020) applied Vision transformer that introduces transformer encoder on the classification of images through computer vision. The transformer model is applied on a sequence of fixed-size image patches that are flattened as vectors. Each of them are linearly embedded, along with position embeddings. The sequence of vectors are fed into the input of the transformer encoder. It learns through embedding where the output of the Transformer encoder is a representation in classification performance. One of the standard approaches on vision transformers is through adding an extra learnable "classification token" to the sequence. Vision transformers had demonstrated effectiveness in Transformer architecture for computer vision tasks.

- **Global Context**

  Extracting the global understanding of an image is the aim of capturing long-range dependency. It provides a good benefit to different recognition tasks on media type classification, object detection and segmentation. (Cao et al., 2019). Introduced by Cao et al. (2019) named the global context provides the benefits of the effective modeling on long-range dependence along with the simplified non-local block with effective modeling with lightweight computation.

It is used by Bouget et al., (2021) on the segmentation of Meningioma on T1-Weighted images. They discussed that global context along with spatial relationships helps models to discriminate between contrast-enhanced meningiomas and bright anatomical structures in the brain. Their study concluded that leveraging the use of global context from a 3D MRI volume provided the best performance even with a limited memory in the GPU.

C. Classification

Leonard (2019) stated that classifying image features is an important research direction in computer vision. CNN can learn image features and classify detected objects through gradient descent methods.

D. Preprocessing

Raw data is usually not ready for correct inferences. Pre-processing is a methodology that converts raw, messy or noisy data to a well organized manner before applying high level machine learning algorithms or any advanced methods for data analysis (Khalid et al., 2020).

- **Skull Stripping**

  Brain extraction or Skull stripping is the process of removing the skull and the background from the MRI image. Smith (2002) created a method named Brain Extraction Tool (BET) that applies a set of locally adaptive model forces to fit the brain's surface. As seen on *Figure 1.3.2*, it is clear from the figure that the skull contains many false positives (Afzal et al., 2022).

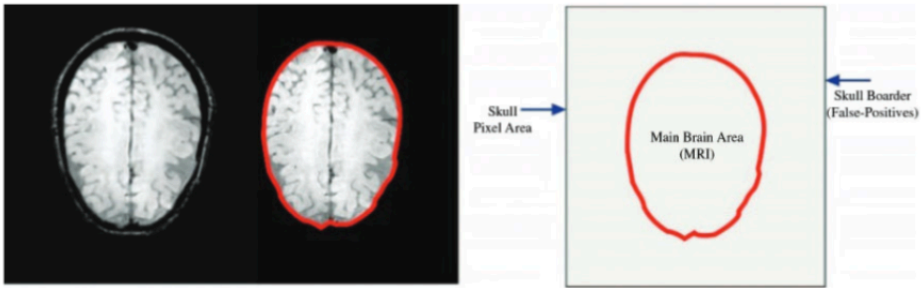


Figure 1.3.2 Skull-part removal from the cerebrum

- **Normalization**

  Medical images tend to be corrupted by noise and artifacts. It is the operation of pixel value rescaling into the range of (0,1) or (-1,1) (Ferlin et al., 2023). Despite of numerous studies have been applied and being effective in this process, this technique can be tedious, and often changes a lot of content in the image (Kociolek et al., 2020)

- **Interpolation**

  To further improve segmentation one of the default operations in preprocessing of medical images is interpolation. The intensity-based registration method is one of the important registration methods (Mahmoudzadeh & Kashou, 2013). It also improves the resolution of images. Mahmoudzadeh & Kashou (2013) concludes that it could play a critical role in the improvement and deterioration of the quality. Assessing the technique is crucial to avoid interpolation error.

D. Conceptual Framework

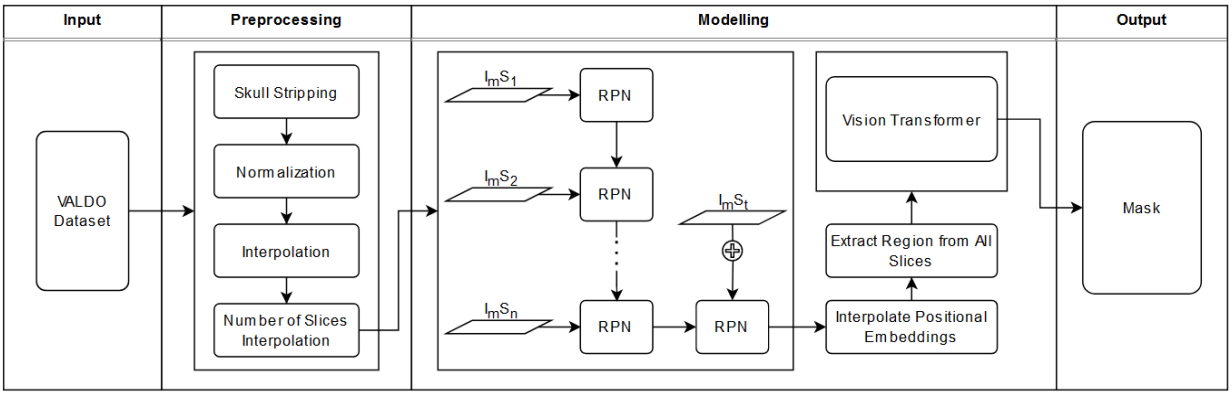


Figure 1.4 Conceptual Framework Diagram

- **Data Preparation**

The VALDO dataset (Sudre et al., 2024) will be utilized for this study, wherein each MRI sample will undergo a series of preprocessing steps for quality inputs in modeling. Initially, each sample will go through skull stripping to remove the skull, background, and non-brain tissues using the method developed by Shattuk et al. (2001), which combines anisotropic diffusion filtering, edge detection, and mathematical morphology. Each sample's local estimates will be computed through a partial volume tissue measurement model, utilizing mean intensity value of the tissue and noise variance values derived from the global image, and categorized into six types of tissues through a maximum posterior classifier based on spatial properties of the brain.

- **Normalization and Interpolation**

After skull stripping, each pixel will undergo normalization, refined by Udupa's approach as modified by Shah et al. (2011). This process undergoes through the creation of a template histogram with region of interest from averaged histograms of a reference population and mapping each subject in a piecewise linear transformation. Wu et al.

(2020) introduced the bilinear interpolation in their intermediate slice synthesis model, which will be used to compute bi-directional spatial transformations and generate the intermediate slice. Particularly, samples have varying slice count that is why another feature will be utilized to set the slice count per sample to adjust the Region Proposal Network inputs accordingly, ensuring consistency in the data.

- **Integrating Global Context with RPNs and Visual Transformer**

The model introduces global context in two ways – through the Region Proposal Network (RPN) and the Visual Transformer (VIT). Each 3D Axial MRI brain scan is broken down slice-by-slice and will be fed to the RPN Transformer. For each subsequent slice in the sample, the RPN utilizes the embedding output from the previous slice to acquire contextual information. After processing all the slices, the target slice will be fed again to the RPN as an input and the final output will be the bounding box vertices for that sample. The decision to utilize the bounding box output from the final slice processed by the RPN ensures that the network incorporates the entire brain's global context before delivering the final output to the ViT.

Once the RPN has identified the bounding box coordinates, the specified region inside those coordinates within each slice is extracted and fed to the Visual Transformer. Normally, a Visual Transformer would take a whole image and separate it into patches to be processed by ViT (Dosovitskiy et al., 2021). However, to integrate global context across axial plane MRI slices, the model will analyze the proposed regions within each slice of the sample. By providing the Visual Transformer consistent patch locations from different slices, the model's attention mechanism integrates a global context to these proposed regions. The transformer will then output a segmentation mask of cerebral

microbleeds at the target slice. This process will continue for each sample, training the model to improve performance.

E. Statement of the Problem

Detecting cerebral microbleeds remains a significant challenge in medical imaging due to the inaccuracies or noise in data, as well as variations in the size and orientation of objects, and differing light conditions (Ferlin et al., 2021). Despite advancements in deep learning, existing methods, such as those developed by Fan et al. (2022) and Chen et al. (2019), still struggle to fully capture contextual information. These methods often rely on a fixed count of slices or the same dimensional size of 3D patch extractions for CMB candidates leading to the loss of crucial information. To address this limitation, Lee et al. (2023) attempted to enhance their model by using MRI scans from multiple planes of the brain, but it resulted in poor performance when using only a single plane, specifically the axial plane.

With that, the main research gap in the literature on automatic CMB detection is the limited research that has considered leveraging the brain's full contextual information within the same plane to enhance the detection capabilities in their models. By focusing on critical areas and applying attention mechanisms, this study hypothesizes that it can improve both the accuracy and reliability of CMB detection. Leveraging the global context of the brain not only resolves the problem of relying on multiple planes but also ensures that crucial surrounding information is not overlooked, thus setting the stage for more accurate identification of microbleeds.

Given the main problem the study's trying to address, it aims to answer the following questions:

1. How do local and global attention mechanisms impact the detection of cerebral microbleeds when evaluated using metrics such as accuracy, precision, sensitivity, and the average number of false positives?
2. What threshold settings should the model learn to determine the optimal size of the Region Proposal Network for accurately detecting cerebral microbleeds?
3. How many adjacent slices should the model consider to effectively reduce false positives in the detection of cerebral microbleeds?

F. Objectives

In light of the problem statement, the study aims to develop a model that utilizes attention mechanisms within a transformer-based region proposal network to improve the detection of cerebral microbleeds. The region proposal network is designed to identify potential cerebral microbleed candidates by concatenating the current slice with the global context and using the Visual Transformer model for the segmentation. Specifically, the study aims to aims to accomplish the following objectives:

1. To determine how local and global attention mechanisms impact the detection of cerebral microbleeds by evaluating metrics such as accuracy, precision, sensitivity, and the average number of false positives.
2. To determine which threshold settings the model should learn to determine the optimal size of Region Proposal Network for accurately detecting cerebral microbleeds.
3. To determine the optimal count of adjacent slices the model should consider to effectively reduce false positives in the detection of cerebral microbleeds.

G. Scope and Limitations

This study mainly investigates the benefits of applying the global context of the whole brain in the segmentation of CMBs from MRI scans, explicitly focusing on images obtained from the axial angle. The study aims to utilize the Transformer model to get the proposed regions from all the slices concatenated with the current slice, essentially bringing the global context of the brain to the second stage. The study will also utilize a Visual Transformer model to get the segmented masks from the proposed regions.

However, the study is limited to the following:

- The study is limited to using T2*-weighted MRI scans from the axial plane.
- The study is limited to the acquired dataset from Where is VALDO? challenge (Sudre et al., 2024).
- The study focuses solely on the segmentation of CMB and will not consider other types of cerebral small vessel disease (CSVD).
- The study does not consider any additional clinical or demographic information about the patients beyond the given MRI scans. The study wants to maintain consistency with the base studies mentioned in this paper..

H. Significance of the Study

The importance of this study lies in its investigation of utilizing a global context to enhance the localized analysis of cerebral microbleeds in MRI images. Previous approaches primarily focused on specific localized regions and neglected broader spatial information. This study aims to bridge this research gap and address previous approaches' limitations in losing crucial information during the detection process. Therefore, the study focuses on improving the accuracy and precision of CMB detection algorithms.

The study may prove to be significant for the following:

- **Healthcare Professionals**. This study aims to provide a tool that accurately detects cerebral microbleeds in MRI images to assist doctors and radiologists in diagnosing and treating brain conditions more efficiently.
- **Patients.** This study aims to give better diagnostic accuracy so patients can receive more reliable results.
- **Medical Researchers.** This study aims to contribute to giving a new perspective on brain imaging. It can serve as a foundation for future research as this new approach in integrating detailed local imaging with the broader context of brain structure could better enhance our ability to diagnose and understand brain conditions.
- **Computer Scientists.** This study aims to showcase an innovative approach for detecting cerebral microbleeds. It demonstrates another approach for developing detection models with the goal of getting high results in accuracy and precision and minimizing the average of false positives. This new approach to integrating global contexts with localizing features could also introduce new research in developing future machine learning models. Lastly, the model's performance in this study will set a benchmark for future research in the application of machine learning in healthcare diagnostics.
- **Research Community.** This study aims to set a benchmark for future research using the VALDO dataset. As one of the first to utilize this open-source dataset and share results publicly, it aims to encourage collaborative research

opportunities. Our published findings can serve as a foundation for others to build upon.

I. Definition of Terms

Presented below are definitions of key terms used throughout this study. This will enhance the readers' comprehension of the research:

- **Accuracy.** It is a metric that measures how often a machine learning model correctly predicts the outcome. It can be calculated by dividing the number of correct predictions by the total number of predictions.
- **Attention Mechanism.** It is a type of neural network architecture with an Encoder-Decoder setup that allows the model to focus on specific input sections while executing a task.
- **Axial Plane.** In terms of basic MRI, it is the view from top to down in the orientation of the human head. It is an X-Y plane parallel to the ground, the head from the feet.
- **Convolutional Neural Networks (CNNs).** It is a neural network commonly used for image classification and computer vision tasks.
- **Coronal plane.** In terms of basic MRI, it is the view from front to back in the orientation of the human head. It is an X-Z plane, the front from the back.
- **Object Detection.** It is the recognition of what objects are inside a given image and where they are.

- **Global Context.** It refers to the whole relationship between all pixels/voxels and the objects. In terms of image processing, it is how every part of an image relates to each other across the entire image, including all slices.
- **Mimics.** In a medical context, it refers to a condition or symptom closely resembling another, potentially leading to a misdiagnosis.
- **Precision.** It indicates a model's performance about the quality of its positive prediction. It is calculated by dividing the number of true positives by the sum of true and false positives.
- **Recurrent Neural Networks (RNNs).** It is a neural network that uses sequential data to solve common temporal issues, especially in language translation and speech recognition.
- **Region Proposal Network (RPN).** It is a fully convolutional network that extracts features of the image. It simultaneously predicts object bounds by learning from feature maps extracted from a base network.
- **Sagittal Plane.** In terms of basic MRI, it is the view from side to side in the orientation of the human head. It is a Y-Z plane, which separates left from right.
- **Sensitivity.** It is a metric used for evaluating how well a model can detect positive instances. It is calculated by dividing the total true positives by the total of true positives and false negatives.
- **Slices.** In terms of MRI, it is the detailed images of the body captured at various depths, allowing the medical experts and researchers to examine the MRI image in different planes.

- **Transformer.** It is a deep learning architecture that is inspired by the multi-head attention mechanism and is commonly used for natural language processing tasks. It is focused on transforming one sequence into another one through encoder and decoder but it does not imply any Recurrent Networks.
- **U-Net.** It is an encoder-decoder convolutional neural network that is commonly used in medical imaging, autonomous driving, and satellite imaging applications.
- **Vision Transformer.** It is a model for image classification that uses a Transformer-like architecture by processing them as a series of patches.
- **Voxel Subspace.** It refers to a specific subset or region within a larger volume represented by voxels in the context of imaging or 3D modeling.

Chapter II Review of Related Literature and Studies

This chapter provides a comprehensive review and analysis of the existing literature and ideas pertinent to this study. Additionally, it justifies the necessity of developing this model by presenting the challenges, constraints, or research gaps found in previous studies on cerebral microbleed detection.

A. Cerebral Microbleeds

Cerebral microbleeds (CMBs) are small circular or elliptical lesions with a typical size of 2 to 10 mm in gradient-recalled echo (GRE) or susceptibility-weighted imaging (SWI) in magnetic resonance imaging (MRI) scans. Histopathologically, CMBs are hemosiderin accumulation in the surrounding vessels and hypointense lesions on brain MRI. Detection of these lesions is essential because these lesions are associated with various cerebrovascular diseases and neurological conditions, including cerebral amyloid angiopathy, hypertensive arteriopathy, and small vessel disease (Lee et al., 2018).

a. Cerebral Microbleeds Mimics

Differentiating CMBs in GRE and SWI is essential because mimics appear similar to CMBS. Common causes of these mimics are:

- **Calcium and Iron deposits**

  These deposits appear similar in CMBs because they appear as small circular low signal intensity on T2*-weighted MRIs typically found bilaterally in the basal ganglia (*Fig. 2.1.1*). Computed tomography (CT) scans can help identify calcification in the choroid plexus, pineal gland, and lobar locations. Although an additional CT scan can help identify

mimics in calcification, this may be considered an inefficient step (Greenberg et al., 2009).

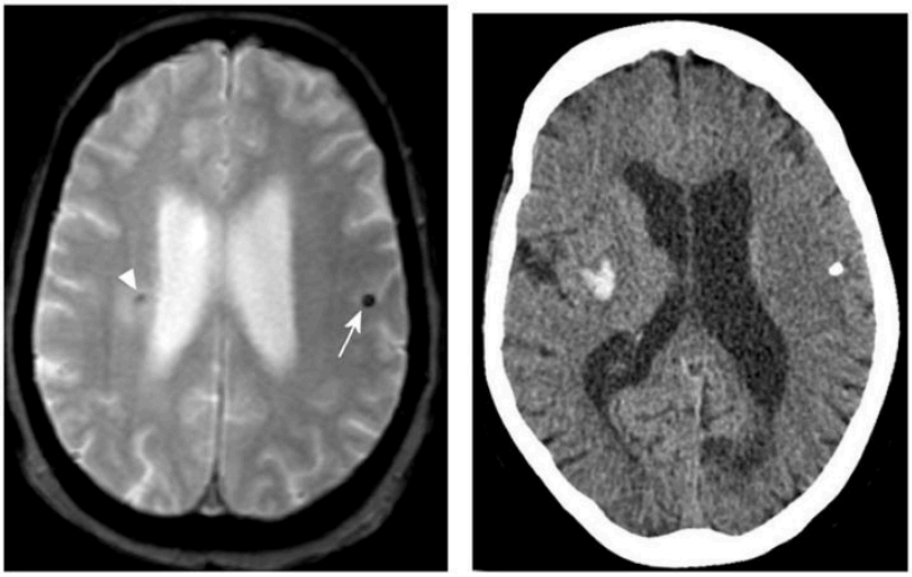


Figure 2.1.1: Left: Axial T2-weighted MRI with calcification in left hemisphere mimicking a CMB (arrow). Right: CT scan showing a high-density lesion and a right hemisphere hemorrhage (arrowhead).

- **Blood flow in pial vessels**

  Flow voids in pial blood vessels in the cross-section in cortical sulci can also mimic CMBs because of their equal visibility on T2-weighted spin-echo and GRE sequences (voids in the artery do not generate a blooming effect). When examined in contiguous slices, these flow voids can be distinguished from CMBs by their sulcal location and linear structure. The presence of paramagnetic deoxyhemoglobin in cerebral venules makes differentiation of CMBs challenging because these also produce their own blooming effect. However, these can be differentiated from CMBs because the reader can see their tubular structure. Differentiating CMB from flow voids is challenging when

CMBs appear adjacent to a small vessel. However, the lesion's termination suggests CMBs are a blind end instead of continuing linearly like a blood vessel as seen in *Figure 2.1.2* (Greenberg et al., 2009).

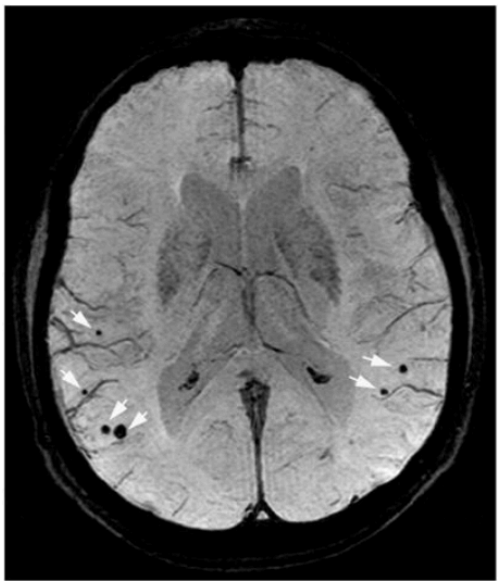


Figure 2.1.2: The image uses a minimum intensity projection of a T2*-weighted MRI (4 mm slab) to show CMBs (highlighted by the arrow) near vessel flow voids on the brain surface.

- **Artifacts from bone structures**

  Artifacts from the bones, especially the artifacts from the air in the sinuses, may obscure CMBs or confuse their interpretation. These usually happen in the temporal and frontal lobes due to the orbit and mastoid bones as seen in *Figure 2.1.3* (Greenberg et al., 2009).

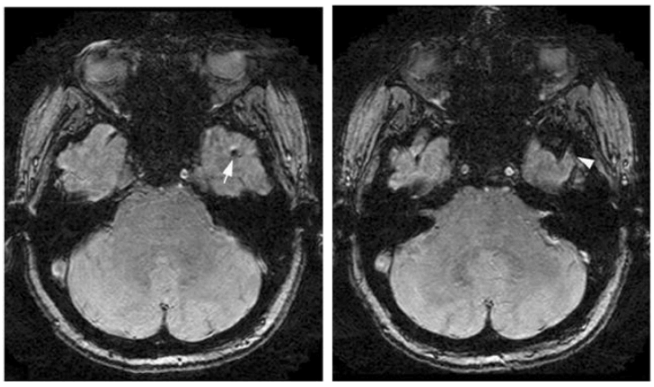


Figure 2.1.3: Axial T2-weighted MRI; left shows potential CMB (arrow), right identifies it as bone artifact (arrowhead).

- **Cavernous malformations**

  Small "type IV" cavernous malformations are considered secondary causes of CMBs. However, when looking at both T1- and T2-weighted MRI sequences, they differ from primary CMBs because they show stagnant blood within the sinusoidal lumen, extravasated blood at various stages of degradation, and a distinctive hemosiderin rim (**Fig. 4**). These malformations can sometimes be confused with the "ringing artifact" seen on T2*-weighted MRI that represents the increased signal within the flow voids (*Figure 2.1.4*) (Greenberg et al., 2009).

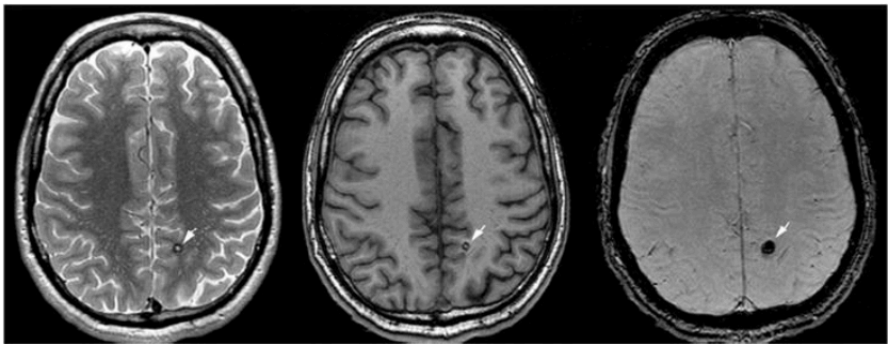


Figure 2.1.4: Axial MRIs (proton density, T1, T2-weighted) showing cavernous malformation as CMB mimic (arrows).

- **Metastatic melanoma**

  This type of melanoma in the brain sometimes appears as hyperintensities on T2*-weighted MRI because of the presence of melanin and the tumors' propensity to bleed. To distinguish from primary CMBs, hyperintensity on T1-weighted MRI is caused by melanin or surrounding edema (following recent intra-tumor hemorrhage). However, small non-edematous lesions can also mimic primary CMBs. (Greenberg et al., 2009)

b. Traditional Ways of Cerebral Microbleed Detection

In clinical practice, the traditional CMB detection method is through visual scoring systems and manual evaluation by radiologists. Microbleed Anatomical Rating Scale (MARS) and the Brain Observer MicroBleed Scale (BOMBS) are examples of visual scoring systems. In both of these methods, they score CMB by the location and by whether the CMBs are certain or ambiguous. The only difference is that the MARS scale divides the cerebral regions into each lobe, and BOMBS divides the cerebral regions into the cortex and subcortical regions. However, manual scoring systems are unreliable without evaluation tools and show insignificant changes even when applied to other MRIs (Lee et al., 2018).

B. Cerebral Microbleed Automatic Detection System

Advancements in deep learning have paved the way for the creation of computer-aided systems in the field of medical imaging. Many studies have specifically designed models that

aim to enhance the detection and analysis of cerebral microbleeds. These microbleeds have significant implications for various neurological diseases, thus making their accurate identification important and critical for effective diagnosis and treatment planning. Leveraging machine learning techniques can help automate and refine the detection process, which can help overcome the limitations of traditional manual methods that are tedious, time-consuming, and subject to human error. The introduction of automatic detection systems for CMBs opens new possibilities for neuro imaging and provides improvements in the field of healthcare.

a. Multiplanar Approach

The multiplanar ensemble model is an approach in medical image analysis that enhances the accuracy and robustness of detection models. Unlike traditional approaches where the model only accepts one plane as an input, the multiplanar ensemble method uses the combined information using ensemble techniques gained from multiple planes simultaneously, mainly orthogonal planes like axial, sagittal, and coronal, to make its predictions.

In the study conducted by Sundaresan et al. (2021), they proposed a Triplanar U-Net ensemble network (TrUE-Net), a method for segmenting white matter hyperintensities (WMH) by using a triplanar ensemble model that combines the predictions of three two-dimensional (2D) U-Net on the axial, sagittal, and coronal planes.

As shown in *Table 2.2.1*, the proposed TrUE-Net scores are among the best in the existing methods for WMH segmentation. This provides evidence that the multiplanar ensemble network does help in the analysis of MRI scans.

Table 2.2.1. *Comparison of existing methods with TrUE-Net*

Method	Type[b]	Population - study[c] (subjects)	Image modalities	SI value	Sens[d]	Spec[e]
Wang et al. (2012)	I,U	Singapore ageing cohort (272)	T1, T2, FLAIR	0.77	0.81	0.97
Gibson et al. (2010)	I,U	WM disease (18)	FLAIR	0.81	-	-
De Boer et al. (2009)	I,S	HC - Rotterdam scan study(6)	T1, PD, FLAIR	0.72	0.79	-
Steenwijk et al. (2013)	I,S	Hypertension (20)	T1, FLAIR	0.84	-	-
Damangir et al. (2012)	IA,S	HC, Alzheimer's, dementia - DemWest (102)	T1, FLAIR	-	0.90	0.99
Ghafoorian et al. (2016)	IA,S	Small vessel disease dataset (50)	PD, T2, FLAIR	-	0.73	-
Yoo et al. (2014)	I,S	Longitudinal/ dementia - Korean study (32)	FLAIR	0.76	-	-
Jeon et al. (2011)	IA,U	SVD - AMPETIS (45)	PD, T2, FLAIR	0.90	-	-
Yang et al. (2010)	IA,U	Dementia - LEILA (30)	FLAIR	0.81	-	-
Shi et al. (2013)	IAAp,U	Acute infarction (91)	DWI, T1, FLAIR	0.84	0.80	-
Samaille et al. (2012)	IAAp,U	MCI, CADASIL (67)	T1, FLAIR	0.72	-	-
Kruggel et al. (2008)	IAAp	HC, dementia - LEILA (116)		-	0.90	0.91
Khademi et al. (2011)	IAAp,U	Subjects with lesions (24)	FLAIR	0.83	0.82	0.99
Admiraal-Behloul et al. (2005)	IA,U	Vasc. disease - PROSPER (100)	PD, T2, FLAIR	0.75	-	-
Anbeek et al. (2004)	IA,S	Arterial vascular disease (20)	T1, IR, PD, T2, FLAIR	0.80	0.97	0.97
Li et al. (2018)	DL	MWSC TS	T1, FLAIR	0.80	-	-
Andermatt et al. (2016)	DL	MWSC TS	T1, FLAIR	0.78	-	-
Berseth, 2017	DL	MWSC TS	T1, FLAIR	0.77	-	-
Valverde et al. (2017)	DL	MWSC TS	T1, FLAIR	0.77	-	-
Kuijf et al. (2019)	DL	MWSC TS	T1, FLAIR	0.77	-	-
Kuijf et al. (2019)	DL	MWSC TS	T1, FLAIR	0.72	-	-
Xu et al. (2017)	DL	MWSC TS	T1, FLAIR	0.73	-	-
Ghafoorian et al. (2017)	DL	Elder SVD (50)	T1, FLAIR	0.78	-	-
BIANCA (Griffanti et al., 2016)	IAAp,S	NDGEN (21)	T1, FLAIR	0.77	0.80	-
		OXVASC (18)	T1, FLAIR, MD	0.74	0.71	-
		MWSC TrS (60)	T1, FLAIR	0.73	0.74	-
TrUE-Net	DL	MWSC TrS (60)	T1, FLAIR	0.91	0.87	-
		NDGEN (20)	T1, FLAIR	0.88	0.86	-
		OXVASC (18)	T1, FLAIR	0.91	0.91	-
TrUE-Net	DL	MWSC TS	T1, FLAIR	0.77	-	-

In the study conducted by Lee et al. (2023), shown in *Figure 2.2.1*, they observed that cerebral microbleeds (CMB) are circularly shaped in the axial plane and elongated elliptical in the sagittal and coronal planes. They have also observed that mimics have a different shape in at least one of the planes. They also mentioned that multiplanar detection is the commonly used method in clinics.

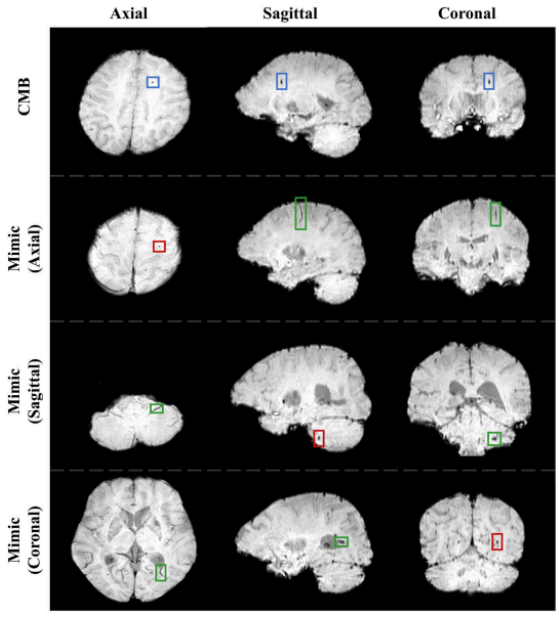


Figure 2.2.1. *Example of CMBs and their mimics from each plane.*

They also identified two significant shortcomings of conventional two-stage models used for CMB detection. The first shortcoming they found is that these models train the models for the first and second stages separately and sequentially. This approach does not pass the extracted features from the first stage to the second stage, and the second stage requires the output from the first stage for training, leading to inefficiencies and potential loss of information. The second shortcoming they found is that current approaches implement three-dimensional (3D) models to minimize the false positives in the second stage only. Although effective in extracting unique features of CMB, significant computational resources are required due to the extensive parameterization involved.

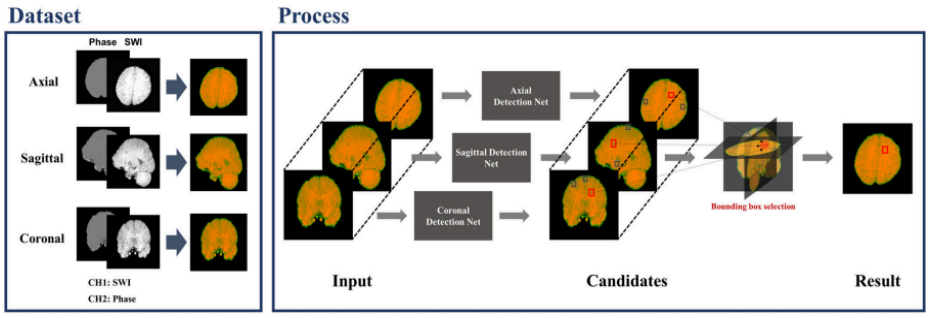
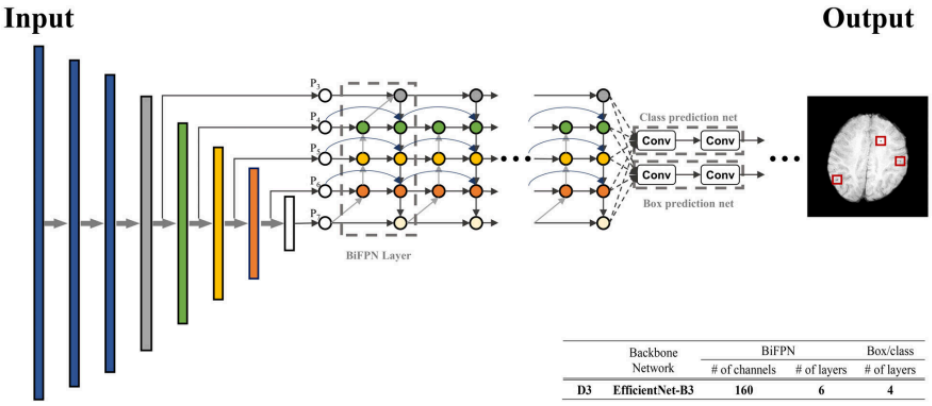


Figure 2.2.2. *The left side illustrates generating input images by combining SWI and phase images. The right side illustrates the ensemble process of the TPE-Det.*



	Backbone Network	BiFPN # of channels	BiFPN # of layers	Box/class # of layers
D3	EfficientNet-B3	160	6	4

Figure 2.2.3. *The architecture of the EfficientDet D3. The table in the right-hand corner is the configurations Lee et al. (2023) used for each plane.*

Lee et al. (2023) proposed an end-to-end model named the Triplanar Ensemble Detection Network (TPE-Det) to overcome these shortcomings. As can be seen in *Figure 2.2.2*, TPE-Det utilizes three 2D detection models based on EfficientDet (the structure for EfficientDet is shown in *Figure 2.2.3*) to individually detect the location of CMBs in each of the three orthogonal planes: axial, sagittal, and coronal. They did not implement classification networks to determine the final candidates of CMBs from the results of each detection model. Instead, they used a bounding box selection to find candidates with morphological characteristics of CMBs from each plane.

With the ensembling of the three planes, TPE-Det had a sensitivity of 96.05% and a precision of 76.76%. *Table 2.2.2* shows the results of the CMB detection and the comparison of various ensemble combinations

Table 2.2.2. *Performance of each ensemble network*

	DS1			
	CMBs Present			CMBs Absent
Ensemble Network	Sensitivity (%)	FP_{avg} (Improvement rate)	Precision (%)	FP_{avg} (Improvement rate)
Axial	97.26	29.43 (baseline)	9.49	30.5 (baseline)
Sagittal	97.26	109.87 (NA)	2.73	117.27 (NA)
Coronal	98.63	261.09 (NA)	1.18	293.08 (NA)
Axial + Coronal	97.26	4.41 (84.88%)	40.49	4.33 (85.80%)
Axial + Sagittal	96.05	2.54 (91.29%)	53.80	2.92 (90.43%)
Axial + Sagittal + Coronal (TPE-Det)	96.05	0.88 (96.99%)	76.76	1.08 (96.45%)

In the study by Fang et al. (2023), they observed that CMB's appearance in the brain is unevenly distributed, with CMBs appearing commonly in specific regions such as the brain lobes and deep brain structures like the basal ganglia,

thalamus, deep white matter, and infratentorial region. With this as the basis, they hypothesized that leveraging the spatial priori knowledge of the whole brain would help improve the accuracy of CMB detection.

Fang et al. also identified several limitations of the current detection methods. Firstly, current 2D and 3D models require high computational costs and a large dataset with proper annotations, which is not practical and hard to obtain. Secondly, 2D analysis of 3D MRI images often suffers from spatial information loss. Thirdly, 2D models have difficulty identifying false positives from the MRI scans. Lastly, current models only consider the appearance of CMBs and not their positional information from the brain.

Fang et al. proposed a Knowledge-guided Body Plane Network (KBPNet) (Fig. 1) to address the current limitations. Their methodology, as the name suggests, is that their model adopts a one-step training and two-step detection framework, where the first step is to train and create the priori knowledge base. The priori knowledge base is a heatmap that represents the risk factors of CMB according to their location.

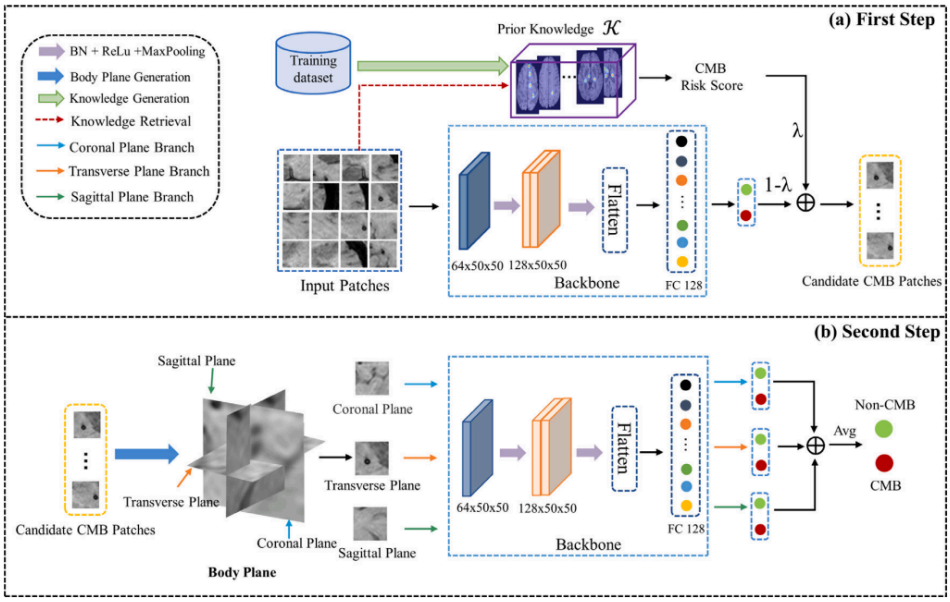


Figure 2.2.4: The overview of the architecture of KBPNet.

The two-step detection model is divided into the preliminary screening of CMB candidates and the fine screening of false positives. In the preliminary screening, they used a 2D CNN classifier model combined with the priori knowledge of CMB distribution to identify candidate CMB patches in the transverse plane. Both image patch information and patch position data are incorporated into the model. The model identifies the candidate patches of CMBs by weighting the CMB risk value and the CMB probability derived from the priori knowledge base and the backbone model. Lastly, in the fine screening, they utilized a 2.5D strategy where they used a blob detection algorithm to localize the CMBs within the patch and center the candidate CMB. These are then mapped into the original 3D coordinates to be fed into the model for further refined classification.

Table 1 shows the comparison of results of the proposed model by Fang et al. with other detection models. It showed that their proposed methodology does help in improving the accuracy of detecting CMBs.

Reference	Method	Layers	Flops	Sensitivity	Precision	FPavg
Chen et al.	3D CNN	–	–	89.10%	–	6.4
Ghafryal et al.	2D CNN	–	–	90.90%	–	4.1
Myung et al.	2D CNN	22	4.06G	66.90%	79.75%	2.15
Al-masni et al.	2D CNN + 3D CNN	11	2869.14M	94.32%	61.94%	**1.42**
Lu et al.	2D CNN	15	172.20M	95.54%	88.37%	5.6
Dadar et al.	2D CNN	50	3.86G	97.37%	83.80%	7.69
Lee et al.	2D CNN	33	2.40G	92.61%	86.74%	3.86
Dou et al.	3D CNN	12	1.12G	93.16%	44.71%	2.74
KBPNet(ours)	**2.5D CNN**	**15**	**706.95M**	**98.24%**	**94.10%**	**1.72**

Table 1: Comparison of KBPNet with other detection models

b. False Positive Reduction Techniques

Cerebral Microbleed mimics are one of the major causes of false positives in automated detection of Cerebral Microbleeds. CMB mimics in SWI or higher field strength MRI become more problematic using methods with higher sensitivity to susceptibility. CMB mimics largely arise in T2*-weighted images to paramagnetic substances other than hemosiderin, including calcifications, iron deposits, or deoxyhemoglobin (Greenberg et al., 2009). The presence of CMB mimics along with algorithmic errors can lead to significant number of false positives (Ferlin et al., 2023). To address these mimics in the differentiation of actual CMBs and their mimics, different approaches were done based on the model's performance and algorithm in detecting the CMB candidate. A comprehensive review of Ferlin et al. (2023) stated that even with the

implementation of this stage, certain approaches still have not managed to acquire satisfying quality.

a. Classifiers

Identified CMB candidates contain features intensity, size, and other complex parameters of a single voxel. These features with fragments of images are to be input on a classifier. Chen et al. (2015) used SVM that learned representation features from the independent subspace analysis (ISA). Along with the candidate screening through intensity values and compact 3D hierarchical feature extraction on the ISA network, it resulted in a high sensitivity rate of 89.44% with an average of 7.7 and 0.9% of false positives per subject and per CMB respectively (Dou et al., 2015).

b. Adjacent slices

Analysis of a sequence of a few adjacent slices based on the features of CMB became essential in the study of Ferlin et al. (2021). They proposed an extra stage where it verifies the ground truth CMB and verification of false-positives as shown in the flowchart of Figure 1. Different configurations of input images are tested to provide information when analyzing the adjacent slices. If the candidate is equal to the features and parameters of the ground truth CMB, therefore the candidate is added to the True positive CMB. This verification was done through Intersection over Union (IOU) of 40%. Specifically, if the bounding box of the candidate's adjacent slices contains 40% of the ground truth CMB bounding box, it is considered as a true CMB. Finally duplicates are

removed (Ferlin et al., 2021). Furthermore Ferlin et al. (2021) have also discussed that enlarging the images and providing information from the adjacent slices by input structuring had made significant improvements on the model's ability to detect CMB. Their study resulted with high precision (89.74%), sensitivity (92.62%), and F1 score (90.84%).

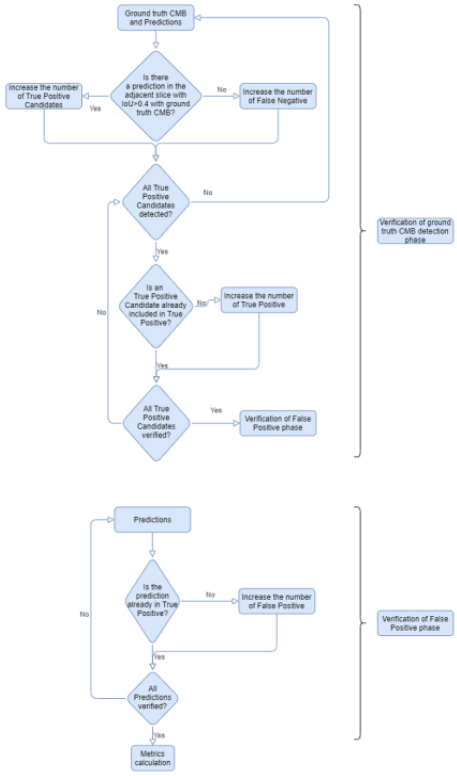


Figure 2.2.4. *Illustrates the post-processing stage of the model by Ferlin et al. (2021)*

c. 2D CNN

The proposed method of Afzal et al. (2021) proposed that after preprocessing the data through skull stripping and augmentation. This was done since they have discussed that the skull area is one of the main causes for higher false positive results. The CMB and CMB mimic images

that were extracted from the SWI samples are augmented to balance the number of each class.

From the K-means clustering stage that provided the division between microbleed and non-microbleed, the proposed approach carried out a false-positive reduction step that is before the classification stage. This is done by removing unwanted connected components that depend on the shape and size of the microbleeds.

Afzal et al. (2021) input the newly false-positive clusters to a transfer-learning classification model. The CNN AlexNet trained on ImageNet data samples was applied that comprises a total of eight layers. This includes convolutional and fully connected layers along with two max-pooling layers and recurrent linear unit (ReLU). ReLu speeds up the training method and stops the overfitting issue in the classification. They also used ResNet 50 to improve accuracy on SWI images. A residual network with 50 layers. The class-specific layers are transferred with the new last layers therefore the entire network was fine tuned for the SWI data samples.

It was discussed based on the findings that the k-means clustering algorithm generated four unique clusters and selected the optimum cluster with a lower False positive. Without preprocessing, Afzal et al. (2021) had changed some steps from the conventional methodology to reduce false positives. The classification stage relies on these methods. The study concluded that ResNet50 and Alexnet along with augmented clustered

data has both 97% accuracy. The utilization of two pre-trained CNN models and a k-means clustering approach to reduce false positives became an effective method to achieve high accuracy in detecting microbleeds.

d. Feed-forward and Feed-Backward feature selection

The goal of the proposed technique of Ghafaryasl et al. (2012) is to distinguish the CMBs from vessels on the base structural analysis in 3D data and to distinguish noisy structures in the image. Similar to the previous methods they include skull-stripping, initial candidate selection, reduction of false-positives using a two layer classification before determining the anatomical location of the CMB seen in Figure 2. Their thesis introduces a semi-automatic detection of cerebral microbleeds in MR images.

Figure 2.2.5. *The generic system model of Ghafarysal et al. (2012)*

They stated that CMBs can be distinguished from vessels through the analysis of their structures in the 3D image. CMBs look like spherical shapes in 3D images while vessels have tubular structures. Noisy structures have plate-like patterns (Ghafarysl et al., 2012). They can also be distinguished based on size and intensity. True CMBs have dark structures with low intensity characteristics in the T2*-MR images. Any candidate without containing these parameters are considered vessels.

Noisy structures have high intensity regions. These features can only be used on a multi-stage system depending on the architecture's design.

In feature extraction, this stage describes how the model distinguishes CMBs from false-positive candidates. They are used as Feed-Forward features selection methods. Through this method they stated that uniform distribution, triangular distribution, beta distribution and gaussian distribution are calculation techniques used in this area. The three key features are extracted on this stage. Geometric-based features contain the size and shape features gained from the binary images. Intensity-based features holds the intensity values of the candidates from the T2*-MRI and Proton Density sequences. Scale-Space representation for feature extraction is a framework that represents the image at multiple scales to find the relevant scale. Through the second order derivative matrix also known as the Hessian matrix, it is used to distinguish the difference of CMB from non-CMB structures. This is used to detect Blob structures in the image. They also used Multiscale vessel enhancement to distinguish spherical structures from tubular or plate-like structures. To detect the strongest blobs and vessel-like structures, they used automatic scale selection.

Partial Least Square Mapping is also applied as a feature extraction. The linear regression model projected both the predicted and observable variables to a new space. The covariance matrix of the predicted variables are also considered.

After extracting the features different classification methods were used to classify the mimics and CMBs. Support vector classifiers that involve optimizing the convex function. The hypothesis dependency on the data through the support vectors makes the model more understandable for the observer. Linear Discriminant Classifier models each class of data by a Gaussian model, Quadratic Discriminant Classifier which the class has there, and Parzen that uses a kernel as a weighting function for estimating the class conditional densities.

To analyze the results Ghafaryasl et al. (2012) used Free receiver Operating Characteristic (FROC). A graph that is used as the criterion for selecting the best performing feature selection method and classifier.

Their study concluded that despite the proposed model containing some False positives, it helps the radiologist to facilitate manual segmentation. It can be used to examine and distinguish CMBs from false positives. The sensitivity for CMB detection was 90% with, on average, 4 false-positives per subject.

C. VALDO Dataset

The “Where is Valdo?” Challenge (Sudre et al., 2024) aims to promote research on automated detection and segmentation for cerebral small vessel disease (CSVD). The University College London and Erasmus MC University Medical Center Rotterdam collaborated and presented the dataset at the satellite event of MICCAI 2021. The dataset consists of medical imaging data obtained from three studies, namely, the Southall and Brent Revisited (SABRE)

(Tillin et al., 2013), Rotterdam Scan Study (RSS) (Ikram et al., 2015), and Alzheimer's and Families (ALFA) (Molinuevo et al., 2016).

The dataset comprises 72 cases with 253 segmented microbleeds. The distribution of cases from the population is 11 cases with 123 segmented microbleeds from SABRE, 34 cases with 96 segmented microbleeds from RSS, and 27 cases with 34 segmented microbleeds from ALFA.

The dataset for microbleeds is composed of data extracted from T1, T2, and T2*-weighted MRIs. Sudre et al. (2024) had expert raters with at least three years of professional experience in medical image analysis to conduct the annotations. ITKSnap segmented the cases from SABER and ALFA and utilized a customized MeVisLab application for the RSS cases. They utilized the BOMBS criteria for the SABRE and ALFA cases and applied the methodology outlined in Vernooij et al. (2008) for the RSS instances. They also ensured the criteria aligned with the STRIVE guidelines (Wardlaw et al., 2013) that show microbleeds as areas with signal voids that are typically 2–5 mm in diameter but can be up to 10 mm.

D. Localized Regions

Deep Neural Network Object localization has gained success in different researches. It has been considered as a requirement before other methods used for medical image analysis. The objective of localization is focused on detecting specific regions of interest (ROI) in medical images(Alaskar et al., 2022). The ROI strategy is to exclude the effects of unnecessary tissues in the MRI to enhance segmentation accuracy (Li et al., 2022). The main difference between local and global components is that global information contains the global shape and structure of the image, local information contains the texture change of the image (Jin et al., 2023).

One of the approaches in using localized regions in MRI images is through the proposal of Yan et al. (2023). Their proposal is applying the idea of Contrastive learning frameworks to be integrated to localized regions. The approach is identifying Super-pixels through Felzenszwalb's algorithm and performs contrastive learning. They enhanced the algorithm that differentiates local regions and formulate the positive and negative pairs. Then perform elastic transformation for both the image and its corresponding local regions to form an augmented image. During implementation all images are re-sampled to have a spacing of 2.5 mm x 1.0 mm x 1.00 mm with respect to the 3D volume's depth, height, and width. Yan et al. (2023) used SGD optimized to pre-train a ResNet-50 encoder for 200 epochs in the global contrast. The whole framework of the local contrast was trained by Adam optimized in an end-to-end fashion.

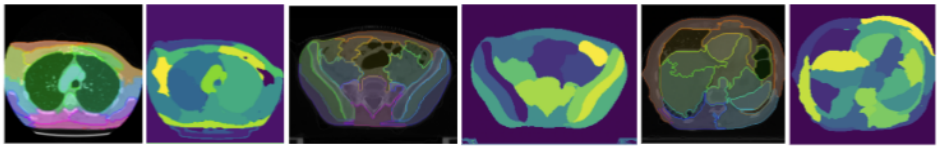


Figure 2.4.1. *The first 3 pictures are generated by Felzenszwalb's algorithm and the embeddings from the local contrast model clustered through K-Means on the right.*

The results show significant improvement, on different datasets. The experiments consistently show LRC boosts the multi-organ segmentation performance of most global contrast models across all three datasets (Yan et al., 2023). They also visualized in *Figure 2.4.1* that is generated by Felzenswalb's algorithm and the features from the LRC. The results demonstrated that LRC learns informative semantic features that can be effectively clustered using a simple K-Means algorithm (Yan et al., 2023). The study concluded that Localized Region Contrasts provides a direction in accurately medical image segmentation improvements.

A study conducted by Chesebro et al. (2021) on applying localization that automatically detects microbleeds in T2*-weighted MRI data. They demonstrated the detection of CMB over

cross-sectional and longitudinal scans ideally for large-scale studies. The algorithm known as MAGIC, identifies CMB through the anatomical localization on brain atlases. Their data consists of T1-weighted images, T2-weighted SWI, and T2-weighted GRE Magnetic resonance images with T1-weighted and T2*-weighted images use the same parameters. The data was visually inspected and determined the microbleeds and its location by three raters (IBM with over eight years of experience, EA and AGC with two years of experience). Fleiss' kappa was used for interrater and intrarater for the assessment of microbleeds (Chesebro et al., 2021).

The data was preprocessed by extracting the brain through FSL Brain Extraction Toolbox of the SWI and GRE T2*-weighted MRI scans. The FSL's MNI lobar mask and T1-weighted images were co-registered to both SWI and GRE images. The CSF mask using the Statistical Parametric Mapping toolbox was used to compute the CSF mask through the co-registered T1-weighted volume. The SWI and GRE scans are then resampled to a higher resolution to a factor of three. Microbleed artifacts would have a diameter of at least six voxels (Chesebro et al., 2021).

The identification of circular regions of interest of potential microbleeds was done by computing each slice of the GRE or SWI image with a 3 x 3 Sobel filter. The edge pixels were then detected through the Canny edge detection algorithm to remove all neighboring voxels that are not local maxima. To remove spurious edges, hysteresis thresholding was used. Hough transform is an algorithm that focuses on a lenient definition of circularity which allows it to be more sensitive to ovoid shapes. The algorithm with a lenient threshold of 80% of the maximum overlap was then used in detecting the regions of interest on all corresponding slices.

The preprocessing methods significantly reduced the potential locations of microbleeds. The remaining overlapping regions of interest were merged together and are too large to be true

microbleeds. These potential microbleeds are excluded and considered as vessels since the lenient cutoff distance between centers is greater than eight pixels (1.15 mm). This method is similar to the visual rating criteria. Finally the region that represents the microbleed candidate was defined as the 3D center and a surrounding neighborhood of a standardized size of 51 x 51 x 25 voxels (two times the maximum expected size of a microbleed). It was used since the neighborhood of this size ensures the candidate is passed for analysis.

Furthermore, to remove false positive locations, Chesebro et al. (2021) used geometric information contained in the regions of interest (ROI). The 3D image entropy, the 2D image entropy of the maximum intensity projection of the ROI and the volume and compactness of the central blob in each ROI as identified through the Frangi filtering. It utilizes the second order derivatives of an image to exact spatial information of the Region's geometry (Chesebro et al., 2021). They also stated that entropy lies on the range from zero to eight in a 8-bit grayscale image. They expected a moderate amount of entropy on 3D images with a large signal void in the center to determine that the ROI is a true microbleed. Higher amount of entropy indicates a noisy false positive region. They also did not expect that 2D entropy data is sharply distinctive as the entropy of 3D data.

The data was again evaluated and corrected by a trained rater after determining the regions of interest. The three raters confirmed that the proposed region was a microbleed. It was identified that 44 microbleed positive participants using SWI scans have 54 locations. 61 locations on the same 44 participants on GRE. MAGIC accurately detected 92% of microbleeds and a reasonable precision seen in *Table 2.4.1*.

***Table 2.4.1**. Detection of CMBs on T2*-weighted MRI Results*

	True positives	False positives	Precision	Average FP/scan
Microbleed positive				
SWI	50	426	0.11	9.7
GRE	56	752	0.07	17.1
Microbleed negative				
SWI	–	249	–	7.32
GRE	–	544	–	15.4
Follow up				
SWI	10	160	0.06	3.64

Inspired from the concept of U-Shape models, Multi-branch learning, and Attention mechanism, Jin et al. (2023) proposed a multi-branch segmentation network that focuses on utilizing local and global information named MBSNet seen in *Figure 2.4.2*. It can be derived from the high-frequency and low-frequency features in images. They also focused on a depth-aware global and local information network. A different approach compared to pixel-level prediction tasks. To obtain global features the image is processed through a dilated convolution and average pooling operation. The global features acts as supplementary information, however it does not combine with local information. The model consists of three stages: local feature extraction branch (Branch L), global feature extraction branch (Branch G), and multi-scale feature fusion branch (Branch F). Taking inspiration from the Inception Mixer, low-level features are used to supplement local information.

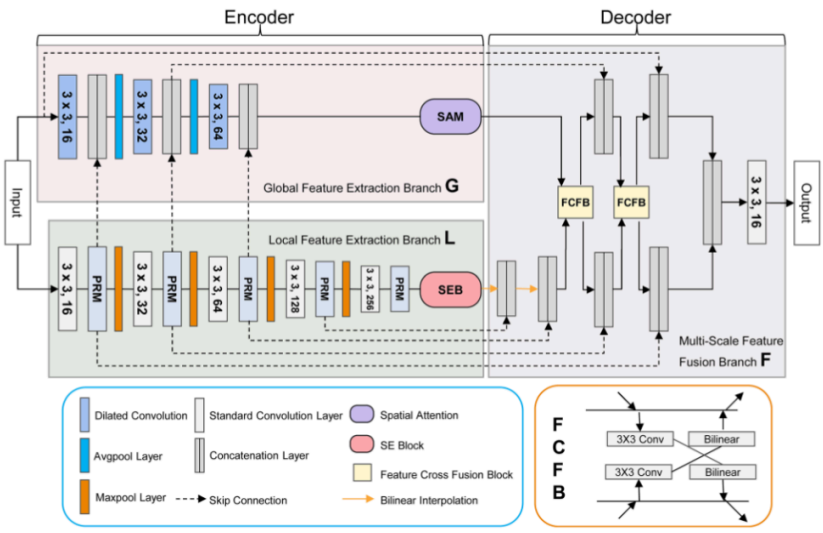


Figure 2.4.2. *The Overview of MBSNet architecture*

The local feature extraction branch is a five-layer convolutional block. The branch L focuses on capturing the local information of the input image. Unlike U-Net, the block of branch L is reduced to 1 convolution, and the channel is reduced by 1/4, respectively 16, 32, 64, 128, 256 (Jin et al., 2023). It retains effective semantic information while saving computing resources and time by reducing the number of parameters. The branch was further enhanced by Parallel residual mixer which consists of two parts: maximum pooling and depth-wise convolutional layer (DWConv) to extract more local information and reduce the feature dimension. Using SEBlock to weigh each channel and guide the learning performance of the model, Branch L can mine different channel characteristics. Further discussing PRM seen in *Figure 2.4.3,* DWConv perceives semantic information details. It also has fewer convolution operation parameters. Normalizing the layer channel through layer normalization and used in recently studied networks the Gaussian error linear unit activation function (GELU) as their activation layer. PRM is denoted as:

Following batch normalization and ReLU activation, f_{1X1} and f_{3X3} are standard convolution layers. $F_{mp}(X)$ is denoted as a max pooling operator $F_{dc}(X)$ is the depth-wise convolution and ⊕ indicating the concatenation.

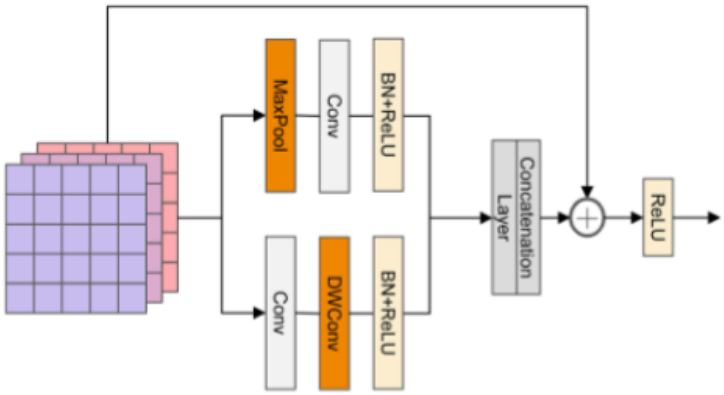


FIgure 2.4.3. *The structure of PRM along with the max pooling operation and depth wise convolutional layer.*

The structure of the dilated convolution block as seen in *Figure 2.4.4*. Global feature extraction branch consists of multiple dilated convolution blocks. Each block has two standard convolution layers and a dilated convolution layer. With three convolution blocks for fine grained pixel-level segmentation tasks, the size of connecting the encoder and the decoder is usually small. This causes information loss which results in down sampling twice. The pooling operation between each convolution block uses average-polling since it retains more complete data. The information is also fused in the down sampling processes. In the formula of Y input feature X in a 1 x 1 convolutional layer f(X) is an inflated convolution that contains BN and ReLU.

To enhance the spatial information of the output Jin et al. (2023) also proposed the spatial attention module (SAM) to branch G integrating the global contextual information. SAM converts the input features into Q, V, K features. The module performs matrix multiplication to

the variables and is added after reshaping. $F_{gl}(X)$ is an adaptive average pooling operator, $F_{sm}(X)$ symbolizes the softmax operator, and $F_{ge}(X)$ is the GELU operator, and $\otimes$ is the multiplication of the matrices.

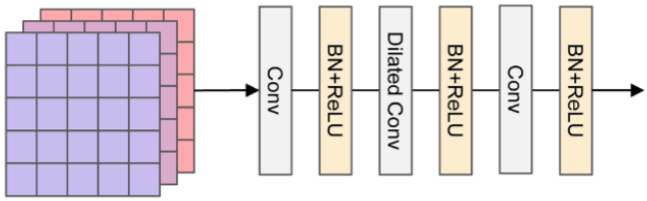


Figure 2.4.4. *The structure of a dilated convolution block extends the field of view through dilated convolutions that perceives global information*

Multi-scale feature fusion branch fuses the global and local information output by the branches L and G. Through bilinear interpolation with skip connections the 4th and 5th layers of L are sampled first. This is enhanced by Feature Cross Fusion Block (FCFB) seen in *Figure 2.4.1*. Its advantages is it can complement features with little calculation. The features are then mapped through an upsampled convolution block containing a standard convolution layer and a bilinear interpolation upsampled layer, respectively.

Through Adam as optimized for the network the experimentation stage was set the initial learning rate to 0.001. Cosine annealing learning rate scheduler with 0.00001 was also applied. The data was augmented through horizontal and vertical flipping, and random cropping. To retain best-performance, Jin et al. (2023) used dice loss for training and Joint loss function of cross entropy. The segmentation model was used on multiple datasets.

MBSNet achieved good results in multiple datasets. It has good results in dermoscopy, gastroscopy, colonoscopy, ultrasound (US), computed tomography (CT), and nuclear magnetic resonance (MRI). It was tested on five datasets ISIC2018, Kvasir, BUSI, COVID-19, and LGG.

On the Kvasir test dataset, it achieved an F1-Score of 85.29%, well above 78.73% for U-Net++ will FLOPs.The proposed model performs better compared to other competitive methods (Jin et al., 2023).

E. Machine Learning Models

Transformers are a new model for deep learning that researchers are utilizing more and more lately in different machine learning problems. It all started with the paper of Vaswani et al., 2017: Attention is All You Need. This paper is the proponent of the architecture encompassing recent deep learning studies. According to Vaswani et al., 2017, the transformer model is a model that creates not only embeddings for the input but positional encoding as well; this makes it great for learning sequences as it provides a context of the input's ordering. However, the main aspect of the transformer model is its self-attention mechanisms. According to Brauwers & Frasincar, 2023, the attention mechanism was first proposed by the study of Daniluk et al., 2017 but popularized by Vaswani et al., 2017 in their paper about transformers. Attention works by having three main general mechanisms: scoring, alignment, and dimensionality (Brauwers & Frasincar, 2023).

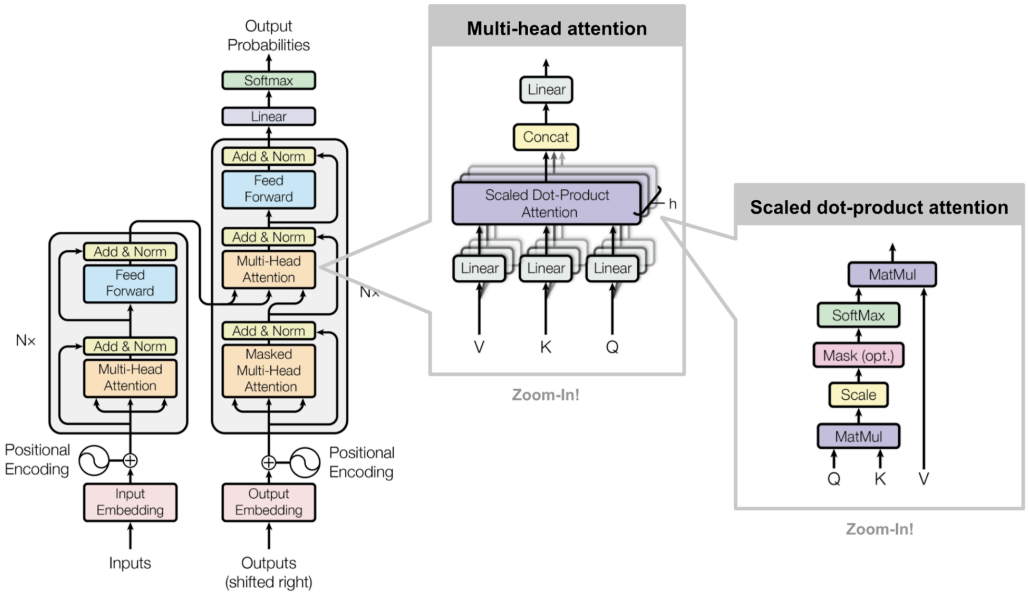


Figure 2.5.1. *Transformer model with more details for multi-head attention*

Since the creation of transformers, more and more studies have been using this architecture and modifying it in some way to work with their goal. In its original use by Vaswani et al., it was for translating English into other languages like French, but transformers can be used for other things such as computer vision. For vision transformers (ViT), images are split into fixed-size patches for linear embedding (including positional encoding), where each of these patches will then be fed to a standard transformer encoder (Dosovitskiy et al., 2021).

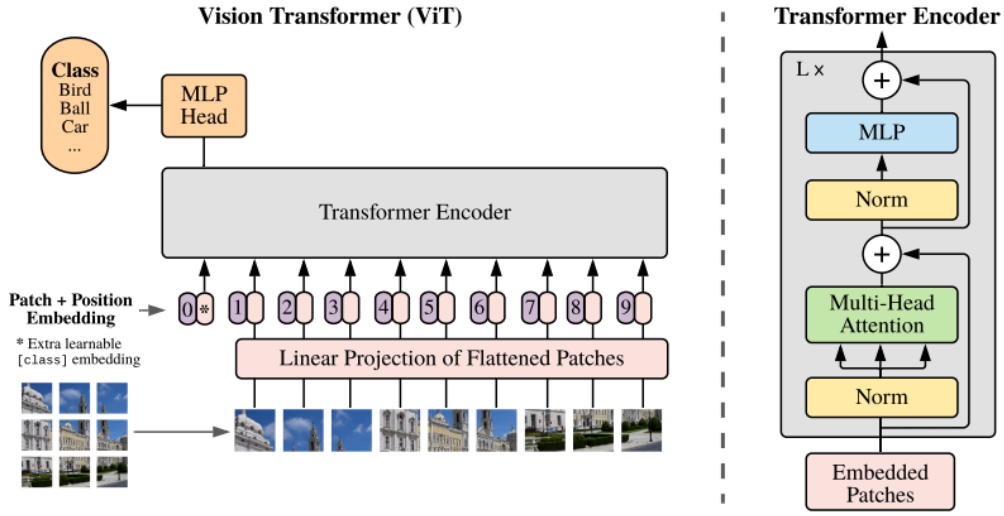


Figure 2.5.2. *The Vision Transformer architecture from Dosovitskiy et al., 2021*

Among all the versions of transformers that have been published, the most consistent aspect that makes these models capable is the attention mechanism within the architectures. By performing a similarity scoring on similar inputs, it can highlight important details that the model should work on.

F. Synthesis

The literature and concepts provided in this chapter have served as a guide for the researchers of this study. The discussed literature has been instrumental in formulating a potential solution to the problem. After a comprehensive review and synthesis of the different literature and concepts, the researchers carefully examined the merits and limitations of various methods, preprocessing techniques, and architectures.

Table 2.6.1. *Cerebral Microbleed Automatic Detection System Synthesis Table*

Title	Author/s	Abstract	Findings
Automatic	Dou, Q. Chen,	Their approach in detecting Cerebral	89.44% of sensitivity

cerebral microbleeds detection from MR images via Independent Subspace Analysis based hierarchical features	H., Yu, L., Shi, L., Wang, D., Mok, V., & Heng, P. (2015)	microbleeds consists of obtaining a region proposed candidate which are screened. Furthermore it extracts 3D hierarchical features and was analyzed to remove false positive candidates through a Support Vector Machine. It supports the idea of how to detect microbleeds and how to eliminate false positives.	was achieved on 19 subjects with 161 CMBs, 7.7 average and 0.9 false positive per subject and per CMB.
A comprehensive analysis of deep neural-based cerebral microbleeds detection system	Ferlin, M., Klawikowska, Z., Grochowski, M., Grzywińska, M., & Szurowska, E. (2023)	Provided a review to all published studies available up until December 2023 that focused on Automated Cerebral Microbleed Detection. Provided an overview on different approaches already done in the study.	Their proposed model achieved with a precision of 89.74%, a sensitivity rate of 92.62%, and a F1 score of 90.84%
A Computer Aided Detection System for Cerebral Micro	Vrooman, H., Lugt, A., & Niessen, W. (2012)	They proposed a CAD system to detect CMBs. Their pre-processing method included skull-stripping. The study focuses on how skull-stripping can be a useful preprocessing method on detecting brain MRI anomalies. Initial candidates will be selected to be followed by reducing false-positives before determining the anatomical location of the CMBs.	The model achieved 90% CMB detection sensitivity with 4 false-positives per subject.
Localized Region Contrast for Enhancing Self-Supervised Learning in Medical Image Segmentation	Yan, X., Naushad, J., You C., Tang H., Sun S., Han, K., Ma H., Duncan J., & Xie X. (2023)	Integrating a Local Region Contrast (LRC) on a proposed novel contrastive learning framework. Their approach identifies Super-pixels through Felzenszwalb's algorithm and performs local contrastive learning. This was done through novel contrastive sampling loss. Their approach provides a background in	LRC has a good direction for accuracy improvement. It is also consistent on multiple datasets.
Automated	Chesbro, A.,	Presented a novel method that	The proposed method

detection of cerebral microbleeds on T2*-weighted MRI	Amarante E., Lao, P., Meier, I., Mayeux, R., & Brickman, A. (2021)	detects microbleeds and identifies over longitudinal scans. It can also identify anatomical localization based on brain atlases.	achieved a sensitivity of 92% with a reasonable precision.
A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning	Jin, S., Yu, S., Peng, J., Wang, H., & Zhao, Y. (2023)	Proposed MBSNet, a novel multi-branch medical image segmentation. Designed with Parallel residual mixer module and dilate convolution block to capture the local and global information of the image. Enhancing the output features through SE-Block and a new spatial attention module. Adopting a cross-fusion method to combine features.	The model achieved an F1-Score of 85.29% above 78.83% from UNet++ on the Kvasir test dataset.
Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images	Vaanathi Sundaresan, Giovanna Zamboni, Peter M. Rothwell, Mark Jenkinson, and Ludovica Griffanti	Proposed a model named TrUE-Net for the segmentation of white matter hyperintensities. The proposed method combines the predictions of each model for each plane of the MRI scans.	The model scored 91% in sensitivity.
Detection of Cerebral Microbleeds in MR Images Using a Single-Stage Triplanar Ensemble Detection Network (TPE-Det)	Haejoon Lee BSc, Jun-Ho Kim BSc, Seul Lee MSc, Kyu-Jin Jung MSc, Woo-Ram Kim ME, Young Noh MD, PhD, Eung Yeop Kim MD, PhD, Koung Mi Kang MD,	They presented a single-stage detection model that uses ensemble techniques to combine the results of three 2D CNNs that are applied to the axial, sagittal, and coronal planes of the MRI.	For the first dataset, the model had 96.05% sensitivity and an average of 0.88 average of FP. For the second dataset, the model had 85.03% sensitivity and adn average of 0.55 FPs.

	PhD, Chul-Ho Sohn MD, PhD, Dong Young Lee MD, PhD, Mohammed A. Al-masni PhD, Dong-Hyun Kim PhD		
Knowledge-guided 2.5D CNN for cerebral microbleeds detection	Zhongding, Rong Zhang, Lijun Guo, Tianxiang Xia, Yingqing Zeng, and Xiping Wu	Proposed a single-step training and two-step detection network called KBPNet. For KBPNet's training, the model uses the information from the distribution of CMBs throughout the whole brain to create the a priori knowledge base. The detection network's first step is to identify the candidates for CMBs and the second stage identifies all the mimics with the help of the priori knowledge base.	The model had 98.24% sensitivity and 94.10% precision, with an average of 1.72 FPs for the first dataset. In the second dataset, the model scored 87.54% sensitivity, 81.63% precision, and an average of 6.54 FPs.

Table 2.6.2. *VALDO Dataset Synthesis Table*

Title	Author/s	Abstract	Findings
Where is VALDO? VAscular Lesions Detection and segmentatiON challenge at MICCAI 2021	Sudre, C. H., Van Wijnen, K., Dubost, F., Adams, H., Atkinson, D., Barkhof, F., Birhanu, M. A., Bron, E. E., Camarasa, R., Chaturvedi, N., Chen, Y., Chen, Z., Chen, S., Dou, Q., Evans, T., Ezhov, I., Gao, H., Girones Sanguesa, M.,	The challenge seeks to enhance the automated detection and segmentation of cerebral small vessel disease (CSVD) imaging markers, which are important for understanding brain h health. These markers include EPVS, CMBs, and lacunes of presumed vascular origin. Twelve teams participated and developed methods to address this challenge using multi-cohort data that is provided.	The challenge demonstrated promising results in the automated detection and segmentation of CSVD markers. This generated methodological research and community engagement toward quantifying CSVD markers in brain MRI scans.

	Gispert, J. D..., de Bruijne, M. (2024)		

Table 2.6.3. *Localized Regions Synthesis Table*

Title	Author/s	Abstract	Findings
-	-	-	-

Table 2.6.4. *Machine Learning Models Synthesis Table*

Title	Author/s	Abstract	Findings
A General Survey on Attention Mechanisms in Deep Learning	Brauwers, G., & Frasincar, F. (2023)	Attention is an important mechanism that can be utilized in different deep learning models. This survey provides an overview of this proposed mechanism in the literature. It also reviews evaluation metrics for attention models and considers future work in this field.	This survey offers a detailed review of attention models. It highlighted their significant performance enhancements in different research fields.
Frustratingly short attention spans in neural language modeling	Daniluk, M., Rocktaschel, T., Welbl, J., & Riedel, S. (2017)	Neural language models have improved prediction of the next token by using attention mechanisms with differentiable memory to capture longer dependencies.	The researchers observed that using an attention mechanism for neural language modeling outperform simpler attention mechanisms. The key-value-predict attention model showed an improvement of 9.4 points over a baseline LSTM.
An image is worth 16x16 words: Transformers	Dosovitskiy, A., Beyer, L., Kolesnikov, A.,	The study demonstrates that the Transformer architecture can effectively handle image classification tasks without relying	The Vision Transformer reached accuracies of 88.55% on ImageNet 90.72%

for image recognition at scale	Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021)	on convolutional neural networks. The Vision Transformer model achieves outstanding results with less training computation when pre-trained on large datasets.	on ImageNet-ReaL, 94.55% on CIFAR-100, and 77.63% on VTAB suite of 19 tasks.
Attention is All you Need	Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017)	Transformer is a new simple network architecture that omits recurrent and convolutional layers and relies on attention mechanisms. This model outperforms existing models on English-to-German and English-to-French machine translation tasks.	The Transformer model achieved a BLEU score of 28.4 in English-to-German translation task and a score of 41.0 in English-to-French task.

Chapter III Research Design and Methodology

A. Hypothesis

This study aims to enhance the detection of cerebral microbleed by leveraging the full contextual information within the same plane through the application of global context. Therefore, this paper hypothesizes the following:

(1) H_0: There is no significant difference in the accuracy, precision, sensitivity, or average number of false positives in the detection of cerebral microbleeds (CMBs) when the researchers utilized local and global attention mechanisms.

(1) H_1: There is a significant difference in the accuracy, precision, sensitivity, or average number of false positives in the detection of cerebral microbleeds (CMBs) when the researchers utilized local and global attention mechanisms.

(2) H_0: There is no significant difference in the accuracy of detecting cerebral microbleeds when the model learns the threshold settings for the optimal size of Region Proposal Network (RPN).

(2) H_1: There is a significant difference in the accuracy of detecting cerebral microbleeds when the model learns the threshold settings for the optimal size of Region Proposal Network (RPN).

(3) H_0: There is no significant difference in the reduction of false positives when the model considers an optimal count of adjacent slices in detection of cerebral microbleeds.

(3) H_1: There is a significant difference in the reduction of false positives when the model considers an optimal count of adjacent slices in detection of cerebral microbleeds.

B. Research Methods

This study will make use of the experimental method, thus defining the problem and hypotheses and identifying a solution to solve the problem. The researchers aim to develop an innovative model that leverages the global contextual information present in MRI scans to enhance the detection of CMBs.

The VALDO dataset obtained from the three studies Southall and Brent Revisited (SABRE) (Tillin et al., 2013), Rotterdam Scan Study (RSS) (Ikram et al., 2015), and Alzheimer's and Families (ALFA) (Molinuevo et al., 2016), will be used in this study. This dataset was specifically chosen because it provides a robust foundation for model training and evaluation.

The model will be evaluated using carefully selected performance metrics, such as accuracy, precision, sensitivity, and the average number of false positives, following the methodology used by Lee et al. (2022). These metrics are important for the comparison of the proposed model against existing models. A t-test will also be applied to assess differences in performance metrics between the proposed and existing models. This will ensure that differences are statistically significant rather than due to random variation. Furthermore, to address any potential biases from dataset imbalances, the study will employ k-fold cross-validation.

C. Research Design

The research will follow these five main phases to achieve the desired output: literature review, data acquisition, preprocessing, modeling, and evaluation.

a. Literature Review

In the literature review phase, a comprehensive collection of related literature, composed of different scholarly articles, studies, and existing research on the detection of cerebral microbleeds, were gathered. The scope of the review extended to the

identification of performance metrics and statistical techniques used to evaluate the current detection model, which is essential for understanding how the researchers have tackled various issues about detecting CMBs, including and not limited to high generation of false positives and low precision rate.

The review of related literature focused on understanding and collecting information about the advancements of detection strategies throughout the years, identifying research gaps, and potential integration of global context and attention mechanisms. This phase serves as the foundation of this study and sets a stage for the development of a novel approach in detecting CMBs while also resolving one of the research gaps in this field of study.

b. Data Acquisition

The selection and acquisition of an appropriate dataset for this study were critical in order to preserve the integrity and success of this study. The researchers found it difficult to find an openly available dataset that is utilized by published papers. After a comprehensive search, the VALDO dataset was found to be an appropriate and reliable dataset for this study. It will be used to train and test the proposed model for detecting CMBs.

c. Preprocessing

The VALDO dataset comprised data gathered from three different studies, resulting in varying characteristics in terms of image resolution, number of slices, slice thickness, quantity, and much more. To achieve uniformity across the dataset, a data cleaning will be applied during the preprocessing phase. This consists of application of

interpolation techniques to standardize the sizes of all samples. Additionally, the researchers will apply stratified k-fold cross validation on the dataset.

d. Modeling

After the preprocessing stage and the cleaning of data, the dataset will be used as the input for the neural network architecture specifically designed to integrate global context and attention mechanisms as discussed in the conceptual framework . This phase will focus on implementing and training the architecture using global attention to detect cerebral microbleeds. It will be in an iterative process of fine-tuning the parameters to achieve the full potential of adding global context in detecting CMBs, thus increasing its accuracy and reliability.

e. Experiment

The objective of the experiment phase is to evaluate the impact of integrating global context into the proposed model compared to a baseline model that lacks these enhancements. Both models will be tested under the same conditions using the same dataset to ensure comparability. In order to evaluate the performance of both models, metrics such as accuracy, precision, sensitivity, and the average number of false positives will be measured.

In this setup, the baseline model will use standard Region Proposal Networks and segmentation models like U-Net as employed in previous studies by Kim et al. (2023) and Fan et al. (2022). The researchers use RPN with a Transformer backbone and Vision Transformers in the segmentation stage. This experimental setup is designed to compare the effects of traditional models versus globally-enhanced models.

f. Evaluation

After training has finished, the model will be tested by the validation set. The results of these tests will be used to evaluate the effectiveness of implementing global attention in detecting CMBs. Performance metrics such as accuracy, precision, sensitivity, and average number of false positives will be used to evaluate this study . Additionally, t-test will be utilized to evaluate the significant difference between the proposed model and the existing models.

D. Research Instruments

a. Hardware

The researchers utilizes personal computers and laptops to conduct this study. The minimum specifications for these devices include 16 GB of RAM, 512 GB of storage, and processors such as but not limited to Intel Core i5 and AMD Ryzen 7. These hardware requirements are important to efficiently handle the computational demands of data processing and model training involved in the study of detecting cerebral microbleeds.

b. Software

This study utilizes Python, one of the most popular programming languages among machine learning professionals because of its easy-to-read syntax, extensive libraries, and cross-platform compatibility. Researchers also employs key libraries including TensorFlow and Keras for neural network modeling and training, NumPy and Pandas for data manipulation, scikit-learn for machine learning algorithm implementation, Matplotlib for data visualization, and NiBabel for neuro imaging data management. During the development and training of models, Google Colab helps the

researchers have a cloud-based platform when using high-performance computing resources. The team uses GitHub for version control and collaboration, Google Docs for documentation, and Jira for project management.

c. Peopleware

Researchers that conduct this study receive guidance and instructions from the thesis coordinator throughout the thesis development process. The thesis adviser provides expertise and knowledge in reigning the paper's content. Furthermore, a radiologist offers invaluable insights into the study of cerebral microbleeds and practical aspects of detecting it, thus enriching the research with clinical expertise.

d. Source of Information

To support the reliability and validity of this study, information was gathered through various academic resources, such as research articles, research publications, and previous theses found on the internet. They navigated platforms such as IEEE, Google Scholar, and ResearchNet to access current and relevant peer-reviewed articles and conference papers which could help the study justify its research methodology. Other sources of information, such as NCBI and ScienceDirect, were also accessed for journals and articles in the fields of neuroscience and medical imaging, which is important to understand the concepts and context of cerebral microbleeds.

E. Sampling and Data Gathering Procedure

The dataset that will be used in this study is the VALDO dataset, which is readily available online. The data collection method primarily involves acquiring the dataset from its hosting platforms. It includes imaging data from multiple cohorts, namely the SABRE, Rotterdam Scan Study (RSS), and Alzheimer's and Families (ALFA). It consists of 72 MRI

scans including T1-weighted, T2-weighted, and T2* sequences, which are important for detecting cerebral microbleeds (Sudre et al., 2024). Specifically, these scans have been used to study cerebral microbleeds across diverse populations and settings, providing a robust dataset for comprehensive analysis.

For the gathered imaging data from the different cohorts, the SABRE cohort provided 11 cases to the training set, RSS contributed 34 cases, and ALFA added 27 cases. The dataset includes segmentations of CMBs across the entire brain for each case. Segmentations are based on a binary classification, where a threshold of 0.5 distinguishes between the presence of a microbleed (1) and the background (0).

There was no need for traditional sampling techniques since the VALDO dataset provides a pre-selected data. This pre-selected data ensures consistency and comparability of the data across various studies, while also meeting the necessary scientific and ethical standards for medical research. Utilizing this dataset aligns with the research objective to enhance the detection of cerebral microbleeds, focusing specifically on analyses using the axial plane.

F. Statistical Treatment of the Data

a. Performance Metrics

The selection of evaluation metrics was critical in this study as these metrics will serve as critical indicators of the performance of the machine learning model in detecting and classifying cerebral microbleeds as true occurrences or mimics. These metrics are key factors in addressing the problem statement and provide valuable data that allow researchers to make analyses and draw well-informed conclusions. To offer a detailed understanding of the model's detection capabilities, this analysis includes counts of true

positives (TP), true negatives (TN), false positives (FP), false negatives (FN), and the average number of false positives.

Accuracy. It measures how often a machine learning model correctly predicts the outcome and is defined by the ratio of the number of correctly predicted observations to the total observations. It answers the question: how often is the model right? This is used for giving a quick overview of the effectiveness of the model.

$$Accuracy\ (\%) = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100$$

Sensitivity (Recall). It measures how well a machine learning model can detect positive instances and is defined by the ratio of actual positives to the sum of actual positives and false negatives. This is particularly important in medical imaging where missing a true case is critical.

$$Sensitivity\ (\%) = \frac{TP}{(TP + FN)} \times 100$$

Precision. It measures how good the quality of a positive prediction made by the model is and is defined by the ratio of correctly predicted positive observations to the total predicted positives. Low precision indicates a high rate of false positives, which is a common challenge in cerebral microbleed detections.

$$Precision\ (\%) = \frac{TP}{(TP + FP)} \times 100$$

Average Number of False Positives. It measures the average number of non-microbleed regions incorrectly predicted as microbleeds per image. It provides direct insight into the challenges of detection models which is a significant concern in this field of study.

$$Average\ Number\ of\ False\ Positives = \frac{total\ FPs\ per\ image}{no.\ of\ images}$$

F1 Score (F-measure). It is calculated as the harmonic mean of precision and sensitivity for a classification model and provides a metric to assess model performance when an uneven class distribution exists. It is useful because it balances the trade-offs between precision and sensitivity.

$$Precision = \frac{(precision \times sensitivity)}{(precision + sensitivity)} \times 2$$

ROC-AUC. It is a performance measurement for classification problems at various threshold settings. It is defined by the area under the curve plotted with the True Positive Rate (TPR) on the y-axis and the False Positive Rate (FPR) on the x-axis as the threshold for classifying a positive prediction. A higher ROC AUC indicates better capability of the model to distinguish between the classes with fewer errors.

$$ROC - AUC = \int_0^1 TPR(FPR^{-1}(x))\, dx$$

Sørensen–Dice similarity coefficient (Dice coefficient). It is used to evaluate the similarity of the predicted positive segment given by the model and the ground-truth positive segment annotated by a clinical expert. It provides a quantitative measure of how closely the predictions of the model aligns with the actual microbleeds determined by radiologists.

$$D = \frac{2 \times TP}{(2 \times TP) + FP + FN}$$

b. Statistical Techniques

K-Fold Cross-Validation. It is widely used in applied machine learning to estimate the skill of machine learning models. This statistical method has a single parameter called k that refers to the number of groups the data sample is split into. The dataset used in this study may contain imbalance data due to the varying prevalence of

cerebral microbleeds across different individuals, which is why this method could help in dealing with the imbalance and provides a more stable estimate of the model's performance. Another limitation of the dataset is that it is limited in size for training and testing, thus this method maximizes the use of available data.

P represents the average performance metric across all foods and P_i is the performance metric (such as accuracy, precision, etc.) of the model on the i-th fold. The formula to calculate the average performance metric P is:

$$P = \frac{1}{k} \sum_{i=1}^{k} P_i$$

T-test. It is a statistical test used to determine whether there is a significant difference between the means of two groups. This will be used to evaluate the performance improvements of the proposed model over existing methods. It will assess whether the observed differences in performance metrics, such as accuracy and precision, are statistically significant.

$\overline{X}_1$ and $\overline{X}_2$ represents the sample means of the two groups, s_1^2 and s_2^2 represents the sample variances, and n_1 and n_2 represents the sample sizes. The formula to calculate the t-test t is:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$