Automatic Detection and Classification of Cerebral Microbleeds Using 3D CNN

M. Mohsin Jadoon [1,2,*], Victor Torres-Lopez [3], Sharjeel A. Butt [2,4], Santosh B. Murthy [5], Guido J. Falcone [3], and Seyedmehdi Payabvash [6]

[1] Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology (PAF-IAST), School of Computing Sciences, Pakistan
[2] Department of Electrical Engineering, International Islamic University Islamabad, Pakistan
[3] Department of Neurology, Yale School of Medicine, New Haven, CT, USA
[4] School of Automation, Central South University, Hunan, China
[5] Department of Neurology, Weill Cornell Medicine, New York, NY, USA
[6] Department of Radiology, Columbia University Medical Center, New York, NY, USA
Email: mohsin.khan@spcai.paf-iast.edu.pk (M.M.J.); victor.torreslopez@yale.edu (V.T.-L.); sharjeel401@csu.edu.cn (S.A.); sam9200@med.cornell.edu (S.B.M.); guido.falcone@yale.edu (G.J.F.); sp4479@columbia.edu (S.P.)
*Corresponding author

***Abstract*—Cerebral Microbleeds (CMBs) are referred to tiny foci of hemorrhage in brain parenchyma which are smaller than 5 (to 10) mm in size. The presence of CMBs is implicated in pathophysiology of cognitive impairment, dementia, radiation-induced vascular injury, traumatic brain injury, hypertensive microangiopathy, and aging. On brain Magnetic Resonance Imaging (MRI) scans, CMBs appear as hypointense foci, most notable on T2*-weighted or Susceptibility-Weighted Imaging (SWI). Detecting these tiny microbleeds with naked eye is a difficult and time-consuming task for radiologists. In this study we developed an algorithm for automatic detection of CMBs. We applied a two-step strategy: at first, we applied pre-processed 2D image dataset to You Only Look Once (YOLO V2) for detection of CMBs. Then, these detected CMBs locations are used to segment 3D patches from their original SWI volume in the datasets. Next, these patches are used as inputs for Convolution Neural Network (CNN). In the second step, we reduced the number of False Positives (FP) and improved our classification accuracy using 3D CNN. We used two datasets consisting of 979 patients: 879 of whom for training of models, and the remainder for independent validation. We were able to achieve an accuracy of 81% and reduce the FP_{avg} to 0.16.**

***Keywords*—Cerebral Microbleeds (CMBs), classification, detection, You Only Look Once (YOLO), 3D Convolution Neural Network (CNN)**

I. Introduction

Cerebral Microbleeds (CMBs) are small foci of hemorrhages that are created by focal accumulations of hemosiderin containing macrophages in brain parenchyma. The paramagnetic properties of these hemorrhagic products lead to susceptibility effects and signal loss on T2*-GRE and Susceptibility-Weighted Imaging (SWI) sequences in brain MRI [1]. The presence of CMBs is associated with higher risk of future intracranial hemorrhage and can be a biomarker for cerebral amyloid angiopathy and cerebrovascular diseases. Recent studies have shown a higher prevalence of CMBs among patients with hematological disorders, brain tumors, abnormalities of blood vessel, hypertension, head trauma, and aneurysm [1]. CMBs are also implicated in pathophysiology of cognitive impairment, and Alzheimer's dementia [2]. On brain MRI scans, CMBs present as tiny black dots which are best seen on three-dimensional T2*-weighted imaging, SWI, and related techniques [3]. Among different MRI sequences [4], SWI series are the most sensitive technique for identification of CMBs [5].

Currently, brain MRI is the most dependable screening modality for identification of CMBs. Utilization of high-field (3T and higher) magnet MRI scanners and sensitive SWI techniques have improved the sensitivity and accuracy of radiologists in detecting tiny CMBs. In current day-to-day clinical practice, radiologists are tasked to identify CMBs, which implies a subjective and tedious process, prone to human errors. Consequently, CMBs may be missed, ignored, or not consistently reported [6]. Identification of CMBs via Computer Assisted Diagnostic (CAD) appears as a viable option to facilitate, expedite, and increase the accuracy of radiologists in detection and quantification of microbleeds.

II. Literature Review

So far, many authors have proposed different automated models for identification of CMBs, many of which limited by small sample size [7]. Barnes *et al.* [8] developed an algorithm with statistical thresholding to recognize

Manuscript received July 23, 2024; revised August 21, 2024; accepted October 16, 2024; published June 12, 2025.

hypointensities inside the images and utilized Support Vector Machines (SVM) to separate confirmed CMBs. They included 126 CMBs in their dataset and achieved a sensitivity of 81.7%. Bian *et al.* [9] proposed a semi-automatic strategy for recognizing CMBs on SWI series. Their algorithm was based on initial radial symmetry transform to detect CMBs followed by exclusion of FPs by using region growing method. A dataset of 15 patients were used in this study. Fazlollahi *et al.* [10] proposed a two-stage model, using multi-scale Laplacian of Gaussian and Random Forests (RF). Their model was validated on 66 patients and achieved 86 % sensitivity. Chen *et al.* [11] developed an algorithm based on CNN. They used a 20-patient dataset with 117 CMBs and achieved a sensitivity of 89.13%. Wang *et al.* [12] also applied CNN with a rank-based pooling scheme to detect CMBs, and achieved 96.94% sensitivity. However, their dataset only included 10 patients. Hong *et al.* [13] described an algorithm based on CNN using transfer learning and ResNet-50, but also using images from 10 patients. Liu *et al.* [14] also used a CNN-based model in a dataset of 195 patients for training and validation, which was tested on images from 25 patients, achieving 95.8% sensitivity. Chen *et al.* [15] used the 3D residual CNN approach, and reduced the FP average results by 89%. Their dataset consists of 73 patients for training and validation; and 12 patients for testing, achieving 90 % sensitivity. Wang *et al.* [16] described a CMB detection method via 2D-DenseNet Neural Network. They also achieved 97% sensitivity, but using a dataset of 10 patients for training and 10 patients for testing. Al-Masni *et al.* [17] exploited a two-stage strategy, where they adopted YOLO and CNN, consecutively, for the detection of CMBs, and reducing FPs. Their dataset consists of 179 patients, including 107 with low resolution and 72 subjects with high resolution images; and they achieved a sensitivity of 78.85% and 93.62%, respectively. Hong *et al.* [18] utilized sliding neighborhood processing and CNN to detect and classify the microbleeds. They achieved 99% sensitivity but only used the data from 10 patients. Doke *et al.* [19] described the Bayesian optimization to find optimum set of parameters for CNN to detect the CMBs. They also generated data from 10 patients using sliding window operation. Lu *et al.* [20] designed a CNN to identify CMBs. Their model achieved average sensitivity of 98.18% by using a dataset of 20 patients. Tummala *et al.* [21] investigated an ensemble of pretrained Vision Transformer (ViT) models (B/16, B/32, L/16, and L/32) for brain tumor classification using T1-weighted MRI images. The ensemble achieved a high test accuracy of 98.7% on a dataset of 3064 MRI slices, demonstrating the potential of ViT models for aiding radiologists in diagnosis. Hossain *et al.* [22] tackles multiclass brain tumor classification using MRI images, evaluating deep learning models such as VGG16, InceptionV3, ResNet50, and others. The proposed transfer learning model, IVX16, achieved the highest accuracy of 96.94 % on a 3264-image dataset. Explainable AI validated the models, and Vision Transformer (ViT) models were compared for performance. Li *et al.* [23] exploited the ground truth for feature enhancement, and then applied these features for training of a CNN. They achieved an average sensitivity and precision of 90% and 79.7%, respectively. Their dataset included 58 patients, with 50 subjects used for training and testing and 8 subjects used for validation.

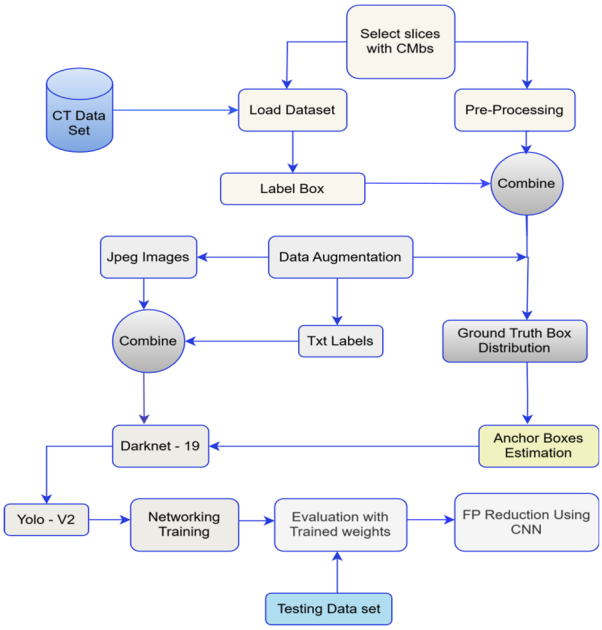


Fig. 1. Flowchart of proposed method.

Over the last decade, Artificial Neural Networks (ANN) could achieve extraordinary milestones in the field of computer vision. Consequently, many researchers have deployed ANN models in the field of biomedical imaging. However, the main obstacle in creating generalizable deep learning models for assessment of clinical images is the relatively small size of samples available for training and validation. This is due to the expensive cost of image acquisition and labeling as well as patients' privacy regulations, which limits public access to medical images. Similarly prior attempts in utilization of deep learning models for detection of CMBs are limited by small sample size in both training and test datasets, as detailed above. Thus, we tried to address this issue by utilizing a large dataset of 979 patients for training and validation of an automated CMB detection model based on YOLO. Fig. 1 describes the flowchart of the proposed method.

The main contributions of this study are:

- Development of a two-step deep learning model for automated CMB detection, utilizing YOLO V2 and Darknet-23 in the first step, followed by a 3D CNN to enhance classification accuracy and reduce False Positives (FPs).
- Training the model on the largest dataset used for CMB detection to date, combining two publicly available datasets, resulting in a total of 879 patients, and testing on 100 patients.

Demonstrating that using a large dataset with preprocessing, augmentation, and fine-tuning significantly improves sensitivity and accuracy while reducing the number of FPs compared to previously reported models.

III. Materials and Methods

A. Preprocessing

Image preprocessing is one of the most important steps in Computer Assisted Diagnostic (CAD) systems. As it is often difficult to extract CMBs due to very homogenous transitions in the MR images, therefore in the model training step we applied a contrast enhancement technique to all those slices that has at least one CMB. i.e.,

$$E \overset{\text{COLFILT}}{\leftarrow} (I, 3)$$

where COLFIT filter enhances the image I depending upon the global mean and global variance of the image [24]. To extend the consistency among the input images intensities values, input slices of subjects were normalized in range 0 to 1.

$$E_{norm} = \frac{E_{orig} - E_{max}}{E_{max} - E_{min}}$$

Here E_{orig}, E_{norm}, E_{min}, and E_{max} refer to pixel intensities. i.e., original, normalized, minimum and maximum respectively.

B. Data Augmentation

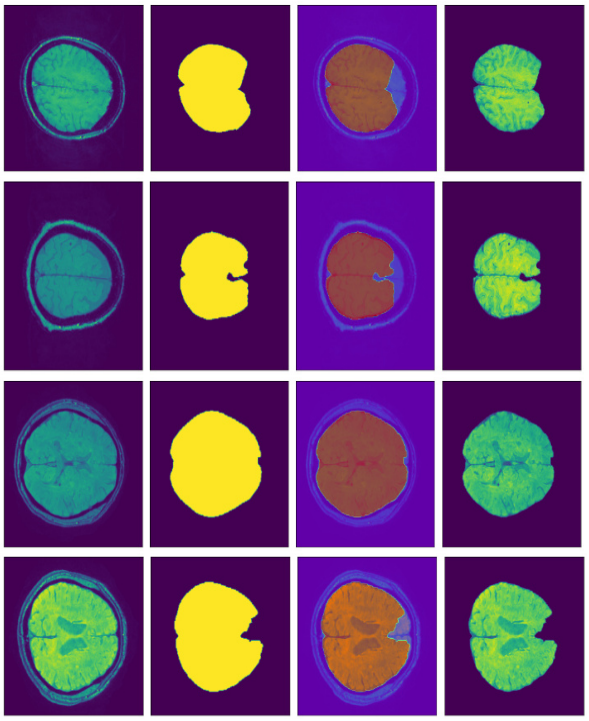
In deep-learning, models are trained to learn a large number of parameters. This will increase the likelihood of over fitting during training due to the model complexity. Data augmentation can artificially increase the number of subjects to alleviate the risk of over fitting [25]. It artificially creates new sample images by applying transformations such as flipping, rotation and other operations to the actual data sample. For every image, we artificially produced seven new sample images using the combination of 0º, 90º, 180º, and 270º rotations and flipping transformations.

C. 2D Slices and Brain Extraction

In the next step, the augmented data is converted into 2D images with size of 448×448. It created an input matrix $\dot{M}$ of size 6096×448×448. Before applying this input image matrix to YOLO, we also removed the brain skull by apply Brain Extraction Tool (BET) on input data [26]. The UK Biobank dataset, however, provides skull-removed SWI series in for their dataset. Fig. 2 contains the examples of original image, binary mask, fused and brain-extracted image. BET operation produces the output images $\dot{N}$ of size 6096×448×448. Corresponding to each image in $\dot{N}$ we have created a text file. These text files and BET output matrix $\dot{N}$ are given as the input to the YOLO.

D. YOLO

YOLO is one of the more recent CNN deep learning techniques which is specialized for object identification in images. It can identify areas of interest in images and characterize their classification. Many other CNNs can also (separately) perform the identification and classification tasks but at much higher computational cost. YOLO can simultaneously perform these two tasks in a single convolutional network as a regression problem and with outputs in form of bounding box and class perdition. Indeed, the region-based YOLO strategy has already been applied in medical images, such as detection of lymphocytes on pathological slides [27], identification of lung cancer on low-dose chest CT [28, 29] and characterization of breast abnormalities on mammograms [29] . An overview of YOLO working is given below [30].



(a) (b) (c) (d)

Fig. 2. Mask results (left to right) (a) Original Image (b) Mask (c) Fused Image (d) Extracted Image.

Mean square error can be computed after computing the loss function for YOLO. Loss function of training, predicting and target of bounding box is given below [30]:

$$Yolo_{loss} = Localization_{loss} + Confidence_{loss} + Classification_{loss}$$

where $Localization_{loss}$ is the error between target and predicted bounding box. $Localization_{loss}$ coefficient can be calculated as:

$$Localization_{loss} = s_1 \sum_{a=0}^{c^2} \sum_{b=0}^{d} l_{ab}^{obj} [(x_a - \ddot{x}_a)^2 + (y_a - \ddot{y}_a)^2] + s_1 \sum_{a=0}^{c^2} \sum_{b=0}^{d} l_{ab}^{obj} \left[\left(\sqrt{w_a} - \sqrt{\ddot{w}_a}\right)^2 + \left(\sqrt{h_a} - \sqrt{\ddot{h}_a}\right)^2 \right]$$

where s_1 is the weight, c belongs to grid cell and d is the count of bounding box in each c. $(x_a - \ddot{x}_a)$ represent the center of c and d, (W, h) express the width and height of d in c. $(\ddot{x}_a, \ddot{y}_a)$ and $(\ddot{w}_a, \ddot{h}_a)$ are the target's center in c. If

here is an object in d in each c then l_{ab}^{obj} is 1 otherwise it is 0.

$Confidence_{loss}$ is the confidence error, if the object is detected in d bounding box of the C. $Confidence_{loss}$ is calculated as under:

$$s_2 \sum_{a=0}^{c^2} \sum_{b=0}^{d} l_{ab}^{obj} \left(conf_a - \ddot{conf}_a\right)^2 + s_3 \sum_{i=0}^{c^2} \sum_{j=0}^{d} l_{ij}^{obj} \left(conf_i - \ddot{conf}_i\right)^2$$

s_2 and s_3 are the weights for the confidence error. $conf_a$ express the confidence score of d in c. $\ddot{conf}_a$ is the confidence score of target's center in c. if there is an object of d in each c, l_{ab}^{obj} is 1 otherwise it is 0.

The $Classification_{loss}$ is the error between conditional probabilities of each class in the grid cell a. It is defined in equation below:

$$s_4 \sum_{a=0}^{c} l_a^{obj} \sum_{\alpha \in \text{classes}} \left(P_a(\alpha) - \ddot{P}_a(\alpha)\right)^2$$

Here s_4, is the weight of Classification error, $P_a(\alpha)$ and $\ddot{P}_a(\alpha)$) represents the probabilities of estimated and actual objects class in cell a.

In this study we used YOLO V2 with darknet-23 for the detection of the CMBs. Fig. 3 explains the basic architecture of YOLO V2. This V2 version of YOLO has better detection results and quick execution due to its high network resolution, multi-level training, anchor boxes and batch normalization. For YOLO V2 training, We have used batch size of 16. Momentum and learning rate are 0.9 and 0.0005 respectively.

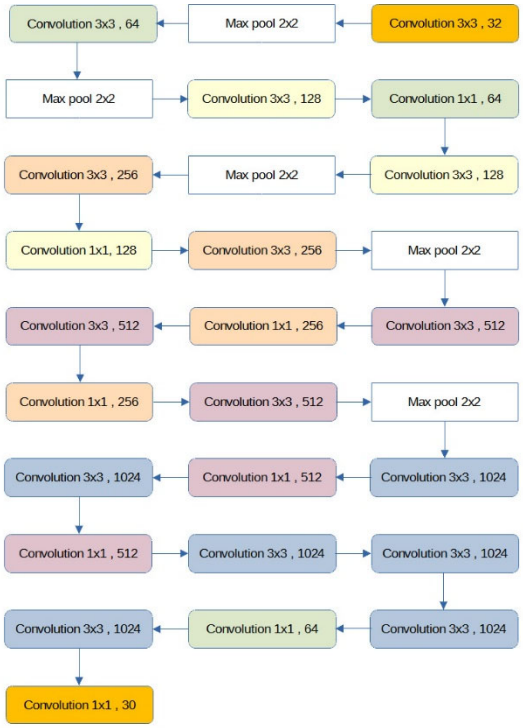


Fig. 3. YOLO V2 architecture.

We applied a two-step strategy: at first, we applied pre-processed 2D image dataset to You Only Look Once (YOLO) for detection of CMBs. Then, these detected CMBs locations are used to segment 3D patches from their original SWI volume in the datasets. Next, these patches are used as inputs for CNN.

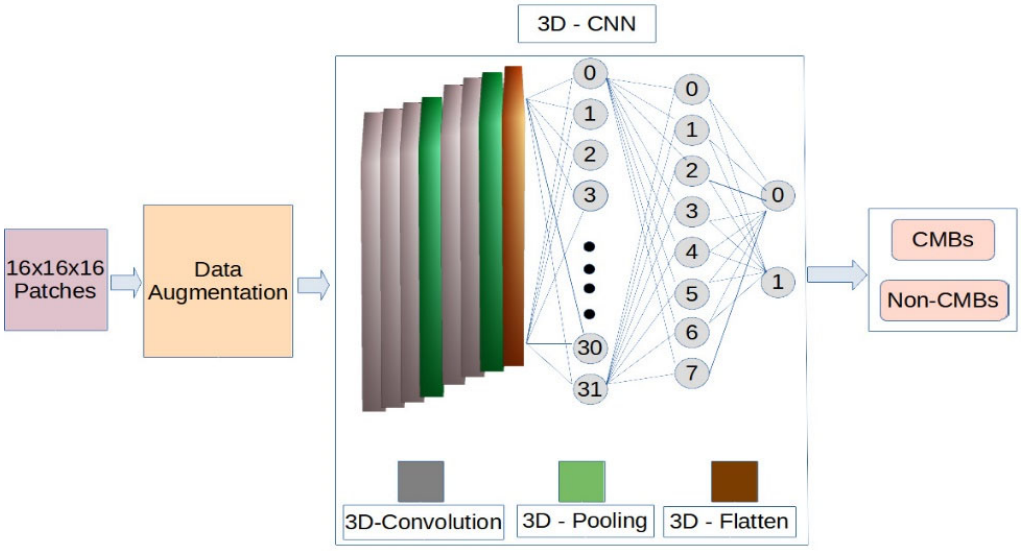


Fig. 4. 3D CNN Architecture.

E. 3D CNN

The proposed 3D CNN architecture is given in Fig. 4. In 3D convolution layer, set of small kernels is convolved with the feature maps of previous layer. The 3D convolution layer can be formulated as below:

$$w_i^l(a,b,c) = f\left(\sum_k \sum_{x,y,z} w_k^{l-1}(a-x, b-y, c-z) h_{ki}^l(x,y,z) + d_i^l\right)$$

Here, w_k^{l-1} represents the K^{th} feature map of $l-1^{th}$ layer. h_{ki}^l is 3D weights for w_i^l and w_k^{l-1} layers. d_i denoted i^{th} biased and $f(.)$ is the activation function.

The activation function used in this model is is $ReLU$ and it is defined as:

$$f(\ddot{x}) = ReLU(\ddot{x}) = \begin{cases} x, & if\ \ddot{x} > 0 \\ 0, & if\ \ddot{x} < 0 \end{cases}$$

In Eq. (7) the output $w_i^l(a,b,c)$ represents the feature maps of the current convolution layer. Maximum polling of 2×2 is used for this study and it can be defined as under:

$$h_{abc,k}^l = \max\left(h_{a'b'c',k'}^{l-1}\right)$$
$$: a \le a' < a+w, b \le b' < b+h, c \le c' < c+d$$

where, $h_{abc,k}^l$ is the output of pooling layer l at abc Position. $h_{a'b'c',k'}$ denotes the 3D cube at region $a'b'c'$ in $l-1^{th}$ layer. Weight, height and depth of pooling layer are represents by w,h,d respectively. Average pooling layer can be defined as given below:

$$h_{abc,k}^l = \frac{1}{w \times h \times d} \sum h_{a'b'c'k'}^{l-1} \begin{matrix} a \le a' < a+w \\ b \le b' < b+h \\ c \le c' < c+d \end{matrix}$$

A flatten layer is used to convert the multidimensional input of previous layer into the one dimensional data. Optimizers are used to calculate and update the network parameters that affect model training and its output. The Adam optimizer have the advantages over other optimizers. It combines the characteristics of AdaGrad and RMSProp optimizer and has high efficiency, convenient implementation and its parameters are updated without gradient transformation. Therefore, for the 3D CNN model we used the Adam optimizer. Adam's update steps are given bellow. To calculate the exponential moving average of the gradient and T_0 is initialized to 0.

$$g_t = \nabla_\theta J(\theta_{t-1}),$$
$$T_t = \alpha_1 T_{t-1} + (1-\alpha_1) g_t$$

Then, calculate the exponential moving average of the square of the gradient; $v_0$is initialized to 0.

$$v_t = \alpha_2 v_{t-1} + (1-\alpha_2) g_t^2$$

The deviation correction is performed on the gradient mean T_t and the gradient variance v_t.

$$\hat{T}_t = \frac{T_t}{(1-\alpha_1^t)}$$
$$\hat{v}_t = \frac{v_t}{(1-\alpha_2^t)}$$

To update the parameters, the initial learning rate γ is multiplied by the ratio of the gradient mean to the square root of the gradient variance.

$$\Theta_t = \theta_{t-1} - \gamma^* \frac{\hat{T}_t}{\sqrt{\hat{v}_t} + \delta}$$

In above equations, α_1 represents the exponential decay rate, which controls the weight assignment, usually taking a value close to 1, with a default of 0.9; α_2 represents the exponential decay rate, which weights the mean of the gradient squares, with a default of 0.99; $\delta = 10^{-7}$, which prevents the denominator from being 0.

Detailed framework of proposed architecture is illustrated in Fig. 5. The five convolution and two max pooling layers are used. Convolution layers have kernels with size 3×3×3 and pooling layers are with kernels of size of 2×2. The dropout ratio used at this stage is 0.4. Whereas, the learning rate is 1×10^{-7} and batch size is 50.

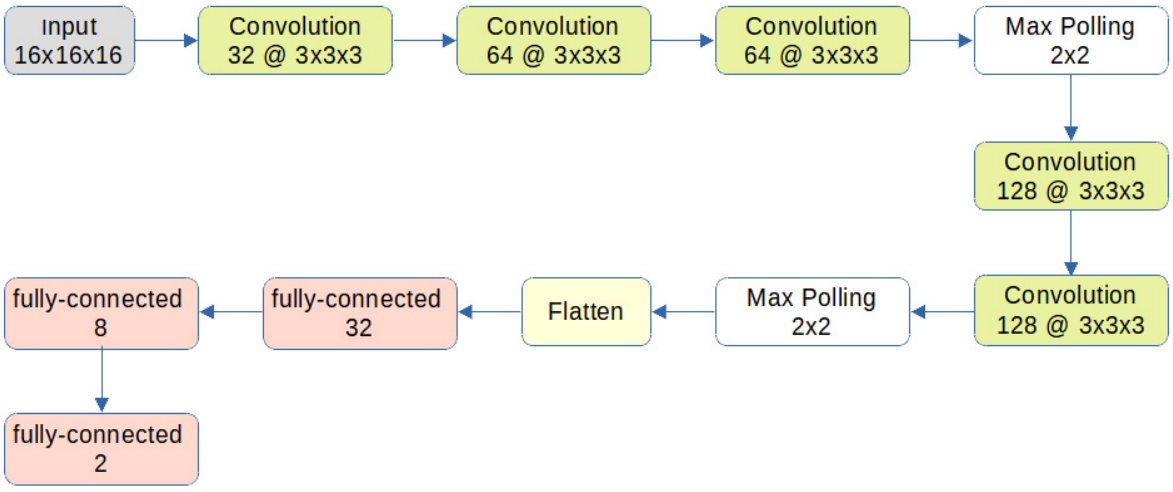


Fig. 5. 3D CNN layers detail.

IV. Datasets

For this study, we utilized two publicly available datasets. The first dataset was from Gachon University Gil Medical Center (GUGMC) [17], and the second dataset was from UK Biobank [31] .

A. Original Dataset

The MRIs in GUGMC dataset were obtained utilizing 3.0 T Verio and Skyra Siemens MRI scanners. It has a total of 179 patients, out of whom, 72 patients had 188 microbleeds with an image matrix size of 512×448×72. The remaining 107 patients had 572 microbleeds and an image matrix of size 288×252×72. We also included 800 patients from the multicentric UK Biobank dataset with an image matrix size of 288×256×48 (https://www.ukbiobank.ac.uk/enable-your-research/about-our-data/imaging-data). All of GUGMC dataset in addition to 100 patients from the UK Biobank were used to train the YOLO model for detection of CMBs.

In FP reduction part, we used all the GUGMC data and 700 patients from UK Biobank to train the 3D-CNN model. For the FP reduction, detected CMBs locations from YOLO were used to segment 16×16×16 patches from their respective images in the dataset. Thus, the size of input image for the 3D-CNN model was set as 16×16×16. Again augmentation operation were applied to these segmented patches. For every image, we artificially produced seven new sample images using the combination of 0º, 90º, 180º and 270º rotations and flipping transformations. Moreover, these rotations and flipping were across each axis. Thus, the resulting dataset contained twenty-two times more images than the original series. The remaining 100 patients from the UK Biobank were used for independent testing of the proposed models.

B. Ground Truth Labeling

The ground truth labels for GUGMC dataset were available online. For UK Biobank dataset, a neuroradiologist (SP)—with 11 years of experience in interpretation of brain MRIs – reviewed and generated the ground truth labeling for the presence of CMBs. The labeling of ground truth was performed as per international standards [32]. The diameter of CMB were set as ≤ 10 mm.

C. Training and Testing

In order to determine the generalizability and reliability of our proposed model, we applied K fold cross-validation, setting the value of K = 5. Both datasets were split into five folds separately, first fold of each dataset were used for validation while other four for training and testing purpose. The proposed research was performed on the clusters of Yale university, with 4 CPU per Node and 3 GPU of NVIDIA-SMI 450.80.02. The CMBs detection and their classification were performed by the Python programming language using keras and tensorflow.

V. Result and Discussion

In this section, results of the proposed algorithms are presented. As we have discussed in literature review section that many studies have adopted two stages strategy for detection of CMBs due to large number of FP in single-stage models. Our method performs well in both detection and FP reduction stages. At Stage 1, it has already achieved very low FP_{avg} and 100% sensitivity even on large test dataset. All other method at Stage 1 neither could able to achieve as low FP_{avg} and nor achieved as high sensitivity and they are missing the patients even in low test dataset.

Although, at Stage 1 our model already has achieved the FP_{avg} of 0.37 that other models not able to achieved even after two stages. The proposed model after utilization of a large dataset for training with appropriate prepossessing, augmentation and fine tuning of the deep learning model have provided higher sensitivity and accuracy as well as lower number of FPs.

A. Evaluation Metrics

The proposed method is evaluated in terms of True Positive (TP), False Positive (FP) patients, specificity, precision, sensitivity, and False Positive average (FP_{avg}).

In medical image classification, a False Positive (FP) is the incorrect classification of subjects, i.e., the model predicts the presence of disease while in reality the subject is disease-negative. On the other hand, a False Negative (FN) is the incorrect classification of subjects where a test result incorrectly indicates the absence of a disease. A TP is the correct classification of positive subjects, whereas a true negative is the correct classification of negative subjects. Specificity is the most commonly used assessment measure, and it represents all the negative cases with TN or FP.

$$Specificity = \frac{TN}{FP + TN}$$

Precision or Positive Predicated Value (PPV) is defined as the number of correctly detected positive cases over all detected positive cases.

$$PPV = \frac{TP}{FP + TP}$$

Sensitivity is defined as the proportion of the detected positive cases over the actual positive cases, including only disease-positive subjects.

$$Sensitivity = \frac{TP}{FN + TP}$$

Accuracy (Acu) is sum of the total number of true values in test dataset and is defined as below:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

The False Positive Average (FP_{avg}) is number of false positive per subject and is defined as:

$$FP_{avg} = \frac{FP}{K}$$

where K represents the number of subjects in testing dataset.

The Matthews Correlation Coefficient (MCC) is used in machine learning as a measure of binary classifications. The coefficient takes into account all the true and false values either they are positives or negatives and is regarded as a balanced measure which can be used even if the classes are of very different sizes. It returns a value between −1 and +1.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

B. Detection of CMBs

First, we trained the model using the GUGMC dataset. This dataset had the 762 CMBs therefore, we gave notation to our trained model as T_{762}. Fig. 6 presents the detection results, including (a) the ground truth labels, (b) the predicted labels, (c) the correctly identified instances (true positives), (d) the incorrectly identified instances (false positives), and (e) the missed detections (false negatives). Table I represents the results of this experiment on validation dataset. With Confidence Score (CS) of 0.5, out of 14 patients 12 of them are detected successfully and 2 patients are detected as FN, the number of FP are 62. Then we evaluated the results with CS of 0.6, where we could only identify 5 patients as TP. There are 9 FN, while 14 are detected as FPs. Given the number of FPs, we trained the model with another 100 patients from UK Biobank. These 100 patients had 102 CMBs so we denoted the model as T_{864}. Again, applying 5-fold cross validation to add additional data in each fold.

Table II describes the results on validation dataset, with 0.5 and 0.6 CS. Here, all 20 patients with CMBs are detected correctly and 37 patients are detected as FP, with sensitivity, specificity, and accuracy of 1, 0.65 and 0.71, respectively. Similarly, with CS of 0.6, model detected 11 patients as TP. There are 15 FP but number of FN increased up to 9.

Thus, the results of the T_{864} model with SC 0.5 were most promising: no FN, and reduced FP rate of 37 in validation dataset. To further improve our model, we trained our model with another set of 100 patients from the UK Biobank. These additional patients had 24 CMBs so we denoted the model as T_{888}. Again, applying 5-fold cross validation, Table III summarizes the results of the model on validation dataset, with 0.5 and 0.6 CS. Model T_{888} successfully detected all patients with CMBs. However, there was an increase in the FP rate. In this model, the CS of 0.6 led to a high rate of FN as well.

As per results summarized in Tables I–III, T_{864} with CS of 0.5 appeared to have the optimal predictions among all models. In Table IV, we present the prediction results in 5 test fold in cross-validation of model T_{864} with 0.5 of CS. In all folds, the models achieved high sensitivity but with high FP rate.

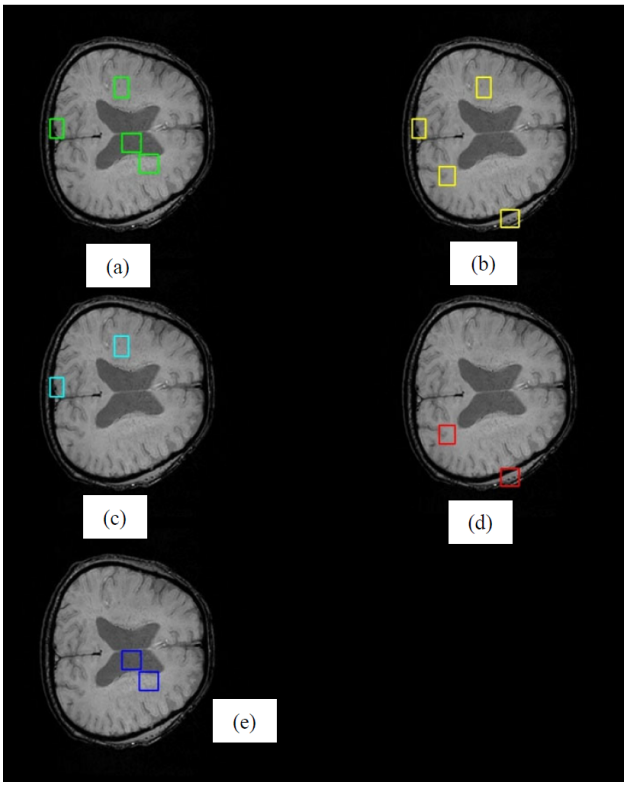


Fig. 6. Detection Results (a) True Labels; (b) Predicted Labels; (c) True Positives; (d) False Positives; (e) False Negatives.

Then, we tested our models on an isolated cohort of 100 patients (Table V). Model T_{762} detects 12 TP with 2 FN, while there are 24 TN and 62 FP. Model T_{864} had better performance, as there is no FN and FP rate is reduced to 37. Model T_{888}, however, only identified 1 TN but with 54 FP.

We also tested our model, applying CS of 0.55, on an isolated cohort of 100 patients (Table VI). Out of 14 patients with CMBs, model T_{762} detects 8 TP with 36 FP. Again, model T_{864} had better performance, but still with 3 FN and 25 FP. Model T_{888} achieved almost same results as T_{864} but with 37 FP.

Table VII summarizes the results of CMB detection applying CS of 0.6 in different models on the 100-patient test dataset. For CMB detection, applying a CS of 0.6 reduced the sensitivity of all models, while increasing their specificity (mostly as a result of the drop in number of FP subjects). With CS of 0.6, low TP rate was the major issue for performance of all models, with model T_{864} again outperforming others.

TABLE I. DETECTION OF CMBS WITH T_{762} ON VALIDATION DATASET

Trained CMBs	Conf score = 0.5			Conf score = 0.6		
T_{762}	TP	FN	FP	TP	FN	FP
	12	2	62	5	9	14

TABLE II. DETECTION OF CMBS WITH T_{864} ON VALIDATION DATASET

Trained CMBs	Conf score = 0.5						Conf score = 0.6					
T_{864}	TP	FN	FP	Sen	Spe	Acu	TP	FN	FP	Sen	Sep	Acu
	20	0	37	1	0.65	0.71	11	9	15	0.55	0.83	0.77

TABLE III. DETECTION OF CMBS WITH T_{888} ON VALIDATION DATASET

Trained CMBs	Conf score = 0.5			Conf score = 0.6		
$\boldsymbol{T_{888}}$	TP	FN	FP	TP	FN	FP
	11	0	54	8	6	17

TABLE IV. FOLD TEST OF MODEL T_{864} WITH CONF SCORE = 0.5

FOLD	TP	FN	FP	TN	Sen	Spe	Acu
Fold 1	11	2	32	61	0.84	0.65	0.67
Fold 2	20	0	37	71	1	0.65	0.71
Fold 3	12	2	37	70	0.85	0.64	0.66
Fold 4	17	0	39	47	1	0.54	0.62
Fold 5	15	1	28	55	0.93	0.66	0.7

TABLE V. DETECTION OF CMBS WITH CONFIDENCE SCORE = 0.5 ON 100 PATIENTS

Trained CMBs	TP	FN	FP	TN	Sen	Spe	Acu
T_{762}	12	2	62	24	0.85	0.27	0.36
T_{864}	14	0	37	49	1	0.56	0.63
T_{888}	13	1	54	32	0.92	0.37	0.45

TABLE VI. DETECTION OF CMBS WITH CONFIDENCE SCORE = 0.55 ON 100 PATIENTS

Trained CMBs	TP	FN	FP	TN	Sen	Spe	Acu
$\boldsymbol{T_{762}}$	8	6	36	50	0.57	0.58	0.58
$\boldsymbol{T_{864}}$	11	3	25	61	0.78	0.7	0.72
$\boldsymbol{T_{888}}$	11	3	37	49	0.78	0.56	0.6

TABLE VII. DETECTION OF CMBS WITH CONFIDENCE SCORE = 0.6 ON 100 PATIENTS

Trained CMBs	TP	FN	FP	TN	Sen	Spe	Acu
$\boldsymbol{T_{762}}$	6	8	14	72	0.42	0.83	0.78
$\boldsymbol{T_{864}}$	10	4	12	74	0.71	0.86	0.84
$\boldsymbol{T_{888}}$	8	6	17	69	0.57	0.8	0.77

C. False Positive Reduction

As depicted in Tables I–VII, all models had relatively high rate of FP in their predictions. To mitigate the high FP rate, we devised a 3D CNN model to improve the classification accuracy.

Table VIII depicts the prediction results from applying 3D CNN on output of YOLO-based models with CS of 0.5. Here, the output of each model (T_{762}, T_{864} and T_{888}) was separately given as input to 3D CNN model for binary classification. Model T_{762} predicted 8 TP and 45 TN, it has sensitivity of 0.57 and FP_{avg} dropped to 0.41. Model T_{888} predicted 11 TP and 54 TN, it has sensitivity of 0.78 and FP_{avg} reduced to 0.32. Model T_{864} produced the best result with 11 TP and 70 TN. There are 3 FN, with FP_{avg} reduced to 0.16. The model achieved an accuracy of 0.81 with sensitivity and specified of 0.78 and 0.81, respectively.

We also tested the results of model after applying CS of 0.55 in YOLO prediction models. Table IX shows that 3D CNN could reduce the FP rate in output of YOLO-based models using CS of 0.55. Model T_{762} predicted 6 TP and 64 TN, with sensitivity of 0.42 and FP_{avg} of 0.22. Model T_{864} predicted 8 TP and 73 TN, achieving specificity of 0.84 with FP_{avg} reduced to 0.13. However, there is a drop in sensitivity to 0.57. Similarly, the 2nd step 3D CNN reduced the FP_{avg} to 0.18 for model T_{888} output, but with a concomitant drop in sensitivity to 0.64.

TABLE VIII. REDUCTION OF FALSE POSITIVE CMBS WITH 3D CNN, CONFIDENCE SCORE = 0.5 ON 100 PATIENTS

Trained CMBs	TP	FN	TN	Sen	Spe	MCC	Acu	$\boldsymbol{FP_{avg}}$
$\boldsymbol{T_{762}}$	8	6	45	0.57	0.52	0.06	0.53	0.41
$\boldsymbol{T_{864}}$	11	3	70	0.78	0.81	0.46	0.81	0.16
$\boldsymbol{T_{888}}$	11	3	54	0.78	0.62	0.28	0.65	0.32

TABLE IX. REDUCTION OF FALSE POSITIVE CMBS WITH 3D CNN, CONFIDENCE SCORE = 0.55 ON 100 PATIENTS

Trained CMBs	TP	FN	TN	Sen	Spe	MCC	Acu	$\boldsymbol{FP_{avg}}$
$\boldsymbol{T_{762}}$	6	8	64	0.42	0.74	0.13	0.7	0.22
$\boldsymbol{T_{864}}$	8	6	73	0.57	0.84	0.35	0.81	0.13
$\boldsymbol{T_{888}}$	9	5	68	0.64	0.79	0.33	0.77	0.18

TABLE X. REDUCTION OF FALSE POSITIVE CMBS WITH 3D CNN, CONFIDENCE SCORE = 0.6 ON 100 PATIENTS

Trained CMBs	TP	FN	TN	Sen	Spe	MCC	Acu	FP_{avg}
T_{762}	4	10	80	0.28	0.93	0.24	0.84	0.06
T_{864}	8	6	78	0.57	0.9	0.34	0.86	0.08
T_{888}	7	7	78	0.5	0.9	0.39	0.85	0.08

TABLE XI. 5 FOLD TEST OF OUR MODEL FOR FALSE POSITIVE REDUCTION TRAINED ON 700, CONF SCORE = 0.5

FOLD	TP	FN	FP	TN	Sen	Spe	Acu
Fold 1	13	2	16	69	0.86	0.81	0.82
Fold 2	18	2	9	71	0.9	0.88	0.89
Fold 3	11	0	12	77	1	0.86	0.88
Fold 4	6	2	4	88	0.75	0.95	0.94
Fold 5	5	2	4	89	0.71	0.95	0.94

TABLE XII. COMPARISON OF EXISTING TECHNIQUES WITH OUR PROPOSED METHOD

Method	Total/Test Patients	Sensitivity	Precision	FP_{avg}
2D-ResNet-50 [13]	10/-	95.71	99.18	3.4
1st stage: 3D-FRST [14]	179/41	99.4	-	276.8
2nd stage: 3D-ResNet		95.24	70.9	1.6
1st stage: 2D-FRST [15]	61/12	86.5	-	231.88
2nd stage: 3D-ResNet		94.69	71.98	11.58
2D-DenseNet [16]	20/-	97.78	97.65	11.8
1st stage: YOLO [17]	151/37	93.62	-	52.18
2nd stage: 3D-CNN		94.32	61.94	1.42
1st stage: 3D-FCN [33]	106/20	98.29	-	282.8
2nd stage: 3D-CNN		93.16	44.31	2.74
Proposed 1st stage: Yolo V2 with T_{864}	279/100	1	-	0.37
Proposed 2nd stage: 3D-CNN CNN with T_{864}	879/100	78.57	40.7	0.16

Table X shows the results from reduction of FP rate after applying 3D CNN on output of YOLO-based models with 0.6 CS threshold. All models achieved >0.8 accuracy and reduced the FP_{avg} to <0.1; however, there is a drop in prediction sensitivity. Models T_{762} , T_{864} and T_{888} achieved sensitivity of 0.28, 0.57, and 0.50 respectively.

In summary, model T_{864} with CS of 0.5 had the best performance after 3D CNN reduction of FP subjects. In Table XI, we present the results of 5 fold test for false positive reduction in model T_{864}. The proposed model was trained on 700 patients dataset. All the CMBs were split into 5 folds. Each time 4 folds were used as training, while 5th fold was used for the validation only. In all 5 fold, model has performed achieved >0.7 sensitivity, >0.8 specificity, and >0.8 Accuracy with lower FP rate compared to the first-step YOLO-based models. Table XII shows the comparison of existing state of art techniques with our proposed method. Our proposed model outperforms the existing techniques, as it has tested on large dataset and able to reduces the FP_{avg} to the significant level.

VI. DISCUSSION

This study proposes a new two-step model for automated detection of CMBs, that is highly generalizable and outperformes previously reported models in terms of sensitivity, accuracy and FP_{avg}. The accurate detection of CMBs is of interest for a plethora of different diseases. The prior research indicates a role of CMBs as diagnostic and prognostic markers for cerebrovascular disease [34]. While CMBs are also found in healthy populations, where prevalence rises strongly with increasing age, they can also be indicative of an underlying small vessel disease [35]. The spatial distribution of the CMBs tends to differ between different causes of small vessel disease, with cerebral amyloid angiopathy most frequently causing lobar CMBs, while classic cerebrovascular risk factors such as hypertension tend to cause deep CMBs [36]. Lobar CMBs were shown to have a high positive predictive value for cerebral amyloid angiopathy, even in patients without manifest Lobar Intracerebral Hemorrhage (ICH) [37]. This suggests that, given correct detection and localization, CMBs could helpful to infer the nature of a present small vessel disease. The overall presence and location of CMBs can also be used as a marker of cerebrovascular risk. The risk of ischemic stroke as well as ICH is increased in stroke-free individuals given the presence of CMBs [38]. CMBs can also play a role when predicting the risk of repeated hemorrhage in patients that initially present with a hemorrhagic stroke [39].

The role of CMBs in cognitive impairment and dementia have also been subject of interest over the past two decades. While the exact causality between loss of cognition and presence of CMBs has not been fully understood yet, two (non-exclusive) hypothesis are commonly discussed: CMBs could have a direct effect onto cognition by disrupting the cerebral network [40], or they could be a manifestation of the underlying brain pathology, which in turn causes the deficits. The

Rotterdam study found an association between presence of microbleeds and decreased cognitive function in patients without dementia, even after adjustment for vascular risk factors and other imaging markers of small vessel disease. Future research on the clinical significance of CMBs as marker of cerebrovascular risk, Alzheimer disease and other areas of interest would greatly benefit from a more sensitive and accurate detection of CMBs.

In this study, we perform different pre-possessing operations, including image enhancement using COLFIT filter is used to enhance the Region of Interest (ROI), image normalization is used to increase the consistency among the input intensities. Moreover, skull stripping is a major phase in MRI brain imaging applications and it refers to the removal of the brain's noncerebral tissues. The main problem in skull-stripping is the segmentation of the non-cerebral and the intracranial tissues due to their homogeneity intensities. In this study, Skull removal was accurately perform using BET. To avoid the over fitting during training due to the model complexity, we used data augmentation. We artificially creates new sample images by applying transformations such as flipping, rotation and other operations to the actual data sample. For CMBs detection we have used YOLO model, finding an optimal model for detection is a challenging problem as the function may have multiple parameters inputs need to be tuned. After several hours long training's and extensive experimental trial optimal selection of model is made. Our proposed model achieved outstanding results at 1st stage. Similarly, for FP reduction a 3D CNN model is proposed, again due to optimal model, architecture layers selection and fine tuning of parameters enable our proposed model to achieved the lowest FP_{avg} rate as compare to other existing models. For medical applications, another main hurdle is creating generalize deep learning models. This is due to the relatively small size of samples available for training and validation. To generalize the results we have used the largest dataset for training and validation for an automated CMB detection model.

VII. Conclusion

New generation of MRI scanners and sequences facilitate depiction of tiny CMBs particularly on SWI series. While CMBs are implicated in several neurological disorders, their identification and reporting in routine clinical practice impose a tedious process for radiologists and is prone to human errors. In this study, we proposed a two-step automated detection algorithm for localization and classification of CMBs. In the first step, we localized CMBs using YOLO V2, which achieved high sensitivity but with relatively high FP rate. Then we used the output of YOLO-based model as an input for a 3D CNN for further improving the classification accuracy and reducing the FP rates. After the FP reduction stage, our proposed model FP rate dropped to 0.16 with 0.81 accuracy, 0.81 specificity, and 0.78 sensitivity. The proposed algorithm can be used for automated detection of CMBs in imaging mega-data to investigate the neurobiological consequences of CMBs in different disease entities.

Conflict of Interest

It is declared that this research work has no conflict of interest.

Author Contributions

The research was conducted by M.M.J., S.A., and S.P., with data labeling undertaken by S.P., S.B.M. and G.J.F. V.T.L., S.B.M., and G.J.F. performed the data analysis, supported by V.T.L. computational expertise. The manuscript was drafted by M.M.J. and S.P., and the final version was approved by all authors.

Funding

This research was supported by the Higher Education Commission (HEC) of Pakistan through the National Research Program for Universities (NRPU) under Grant Number 15212/NRPU/R&D/HEC/2021, as well as a travel grant for presentation at the 2024 International Conference on Biomedical Signal Processing (ICBSP) in Hong Kong. S.P. is supported by Doris Duke Charitable Foundation (2020097) and NIH (K23NS118056).

Acknowledgment

We are thankful to Gil Medical Center, Gachon University Korea, AND UK Biobank for making their dataset publicly available.

References

[1] S. Martinez-Ramirez, S. M. Greenberg, and A. Viswanathan, "Cerebral microbleeds: overview and implications in cognitive impairment," *Alzheimers Res. Ther.*, vol. 6, pp. 1–7, 2014.

[2] H.-C. Koennecke, "Cerebral microbleeds on MRI: Prevalence, associations, and potential clinical implications," *Neurology*, vol. 66, no. 2, pp. 165–171, 2006.

[3] F. Bonneville, F. Cattin, K. Marsot-Dupuch, D. Dormont, J.-F. Bonneville, and J. Chiras, "T1 signal hyperintensity in the sellar region: spectrum of findings," *Radiographics*, vol. 26, no. 1, pp. 93–113, 2006.

[4] M. W. Vernooij, M. A. Ikram, P. A. Wielopolski, G. P. Krestin, M. M. B. Breteler, and A. van der Lugt, "Cerebral microbleeds: accelerated 3D T2*-weighted GRE MR imaging versus conventional 2D T2*-weighted GRE MR imaging for detection," *Radiology*, vol. 248, no. 1, pp. 272–277, 2008.

[5] E. M. Haacke, Y. Xu, Y. N. Cheng, and J. R. Reichenbach, "Susceptibility Weighted Imaging (SWI)," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 52, no. 3, pp. 612–618, 2004.

[6] D. Dao, A. N. Fraser, J. Hung, V. Ljosa, S. Singh, and A. E. Carpenter, "CellProfiler analyst: Interactive data exploration, analysis and classification of large biological image sets," *Bioinformatics*, vol. 32, no. 20, pp. 3210–3212, 2016.

[7] M. A. Morrison *et al.*, "A user-guided tool for semi-automated cerebral microbleed detection and volume segmentation: Evaluating vascular injury and data labelling for machine learning," *Neuroimage Clin*, vol. 20, pp. 498–505, 2018.

[8] S. R. S. Barnes, E. M. Haacke, M. Ayaz, A. S. Boikov, W. Kirsch, and D. Kido, "Semiautomated detection of cerebral microbleeds in magnetic resonance images," *Magn Reason. Imaging*, vol. 29, no. 6, pp. 844–852, 2011.

[9] W. Bian, C. P. Hess, S. M. Chang, S. J. Nelson, and J. M. Lupo, "Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images," *Neuroimage Clin*, vol. 2, pp. 282–290, 2013.

[10] A. Fazlollahi *et al.*, "Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging," *Computerized Medical Imaging And Graphics*, vol. 46, pp. 269–276, 2015.
[11] H. Chen, L. Yu, Q. Dou, L. Shi, V. C. T. Mok, and P. A. Heng, "Automatic detection of cerebral microbleeds via deep learning based 3D feature representation," in *Proc. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2015, pp. 764–767.
[12] S. Wang, Y. Jiang, X. Hou, H. Cheng, and S. Du, "Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling," *IEEE Access*, vol. 5, pp. 16576–16583, 2017.
[13] J. Hong, H. Cheng, Y.-D. Zhang, and J. Liu, "Detecting cerebral microbleeds with transfer learning," *Mach Vis Appl*, vol. 30, no. 7, pp. 1123–1133, 2019.
[14] S. Liu *et al.*, "Cerebral microbleed detection using susceptibility weighted imaging and deep learning," *Neuroimage*, vol. 198, pp. 271–282, 2019.
[15] Y. Chen, J. E. Villanueva-Meyer, M. A. Morrison, and J. M. Lupo, "Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network," *J. Digit. Imaging*, vol. 32, pp. 766–772, 2019.
[16] S. Wang, C. Tang, J. Sun, and Y. Zhang, "Cerebral micro-bleeding detection based on densely connected neural network," *Front Neurosci.*, vol. 13, 422, 2019.
[17] M. A. Al-Masni, W.-R. Kim, E. Y. Kim, Y. Noh, and D.-H. Kim, "Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach," *Neuroimage Clin*, vol. 28, 102464, 2020.
[18] J. Hong, S.-H. Wang, H. Cheng, and J. Liu, "Classification of cerebral microbleeds based on fully-optimized convolutional neural network," *Multimed Tools Appl*, vol. 79, no. 21, pp. 15151–15169, 2020.
[19] P. Doke, D. Shrivastava, C. Pan, Q. Zhou, and Y.-D. Zhang, "Using CNN with Bayesian optimization to identify cerebral micro-bleeds," *Mach. Vis. Appl.*, vol. 31, pp. 1–14, 2020.
[20] Z. Lu, Y. Yan, and S.-H. Wang, "CMB-net: A deep convolutional neural network for diagnosis of cerebral microbleeds," *Multimed. Tools Appl.*, vol. 81, no. 14, pp. 19195–19214, 2022.
[21] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling," *Current Oncology*, vol. 29, no. 10, pp. 7498–7511, 2022.
[22] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification," *IEEE J. Biomed Health Inform*, vol. 28, no. 3, pp. 1261–1272, 2023.
[23] T. Li *et al.*, "Detecting cerebral microbleeds via deep learning with features enhancement by reusing ground truth," *Comput. Methods Programs Biomed*, vol. 204, 106051, 2021.
[24] R. C. Gonzalez, *Digital Image Processing*, Pearson Education India, 2009.
[25] T. Deserno, M. Soiron, J. Oliveira, and A. Araujo, "Towards computer-aided diagnostics of screening mammography using content-based image retrieval," in *Proc. 2011 24th SIBGRAPI Conference on Graphics, Patterns and Images*, IEEE, 2011, pp. 211–219.
[26] S. M. Smith, "Fast robust automated brain extraction," *Hum Brain Mapp.*, vol. 17, no. 3, pp. 143–155, 2002.
[27] M. van Rijthoven, Z. Swiderska-Chadaj, K. Seeliger, J. van der Laak, and F. Ciompi, "You only look on lymphocytes once," *Medical Imaging with Deep Learning*, 2018.
[28] J. George, S. Skaria, and V. V Varun, "Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans," in *Proc. Medical Imaging 2018: Computer-Aided Diagnosis*, SPIE, 2018, pp. 347–355.
[29] M. A. Al-Masni *et al.*, "Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system," *Comput Methods Programs Biomed*, vol. 157, pp. 85–94, 2018.
[30] J. Redmon, "You only look once: Unified, real-time object detection," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
[31] K. L. Miller *et al.*, "Multimodal population brain imaging in the UK Biobank prospective epidemiological study," *Nat Neurosci*, vol. 19, no. 11, pp. 1523–1536, 2016.
[32] S. M. Greenberg *et al.*, "Cerebral microbleeds: a guide to detection and interpretation," *Lancet Neurol*, vol. 8, no. 2, pp. 165–174, 2009.
[33] Q. Dou *et al.*, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1182–1195, 2016.
[34] L. Puy *et al.*, "Cerebral microbleeds: from depiction to interpretation," *J. Neurol Neurosurg Psychiatry*, vol. 92, no. 6, pp. 598–607, 2021.
[35] M. M. F. Poels *et al.*, "Prevalence and risk factors of cerebral microbleeds: an update of the Rotterdam scan study," *Stroke*, vol. 41, no. 10,_suppl._1, pp. S103–S106, 2010.
[36] A. Charidimou, L. Pantoni, and S. Love, "The concept of sporadic cerebral small vessel disease: A road map on key definitions and current concepts," *International Journal of Stroke*, vol. 11, no. 1, pp. 6–18, 2016.
[37] S. Martinez-Ramirez *et al.*, "Diagnostic value of lobar microbleeds in individuals without intracerebral hemorrhage," *Alzheimer's & Dementia*, vol. 11, no. 12, pp. 1480–1488, 2015.
[38] A. Charidimou *et al.*, "Clinical significance of cerebral microbleeds on MRI: A comprehensive meta-analysis of risk of intracerebral hemorrhage, ischemic stroke, mortality, and dementia in cohort studies (v1)," *International Journal of Stroke*, vol. 13, no. 5, pp. 454–468, 2018.
[39] A. Charidimou *et al.*, "Brain hemorrhage recurrence, small vessel disease type, and cerebral microbleeds: A meta-analysis," *Neurology*, vol. 89, no. 8, pp. 820–829, 2017.
[40] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.