

YOLOv5s-CAM: A Deep Learning Model for Automated Detection and Classification for Types of Intracranial Hematoma in CT Images

*Original*

YOLOv5s-CAM: A Deep Learning Model for Automated Detection and Classification for Types of Intracranial Hematoma in CT Images / Vidhya, V.; Raghavendra, U.; Gudigar, Anjan; Basak, Sudipta; Mallappa, Sankalp; Hegde, Ajay; Menon, Girish R.; Barua, Prabal Datta; Salvi, Massimo; Ciaccio, Edward J.; Molinari, Filippo; Acharya, U. Rajendra. - In: IEEE ACCESS. - ISSN 2169-3536. - ELETTRONICO. - 11:(2023), pp. 141309-141328. [10.1109/ACCESS.2023.3339560]

*Availability:*

This version is available at: 11583/2984631 since: 2023-12-20T12:48:42Z

*Publisher:*

IEEE

*Published*

DOI:10.1109/ACCESS.2023.3339560

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

Received 10 November 2023, accepted 28 November 2023, date of publication 5 December 2023,  
date of current version 19 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3339560

## RESEARCH ARTICLE

# YOLOv5s-CAM: A Deep Learning Model for Automated Detection and Classification for Types of Intracranial Hematoma in CT Images

V. VIDHYA<sup>1</sup>, U. RAGHAVENDRA<sup>2</sup>, ANJAN GUDIGAR<sup>2</sup>, SUDIPTA BASAK<sup>2</sup>,  
SANKALP MALLAPPA<sup>2</sup>, AJAY HEGDE<sup>3</sup>, GIRISH R. MENON<sup>4</sup>, PRABAL DATTA BARUA<sup>5,6,7</sup>,  
MASSIMO SALVI<sup>8,9</sup>, EDWARD J. CIACCIO<sup>10</sup>, FILIPPO MOLINARI<sup>8,9</sup>, (Senior Member, IEEE),  
AND U. RAJENDRA ACHARYA<sup>11,12</sup>, (Senior Member, IEEE)

<sup>1</sup>Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

<sup>2</sup>Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

<sup>3</sup>Manipal Hospitals, Bengaluru, Karnataka 560102, India

<sup>4</sup>Department of Neurosurgery, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

<sup>5</sup>Cogninet Brain Team, Cogninet Australia, Sydney, NSW 2010, Australia

<sup>6</sup>School of Business (Information Systems), Faculty of Business, Education, Law & Arts, University of Southern Queensland, Toowoomba, QLD 4350, Australia

<sup>7</sup>Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

<sup>8</sup>Biolab, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy

<sup>9</sup>PolitoBIOMed Laboratory, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy

<sup>10</sup>Department of Medicine, Columbia University, New York City, NY 10032, USA

<sup>11</sup>School of Mathematics, Physics, and Computing, University of Southern Queensland, Springfield, QLD 4300, Australia

<sup>12</sup>Centre for Health Research, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Corresponding author: Anjan Gudigar (anjan.gudigar@manipal.edu)

This work was supported by the Open Access funding provided by the Manipal Academy of Higher Education, Manipal.

**ABSTRACT** Intracranial hematoma due to traumatic brain injury is a serious health concern with rates of morbidity and mortality that are increasing worldwide. Manual identification is slow, subject to observer variabilities, and the existing automated techniques for intracranial hematoma detection in non-contrast computed tomography images cannot effectively detect multiple lesions of irregular sizes and shapes. Therefore, a computer-aided system using different window settings, YOLOv5s, cascaded attention module, and spatial pyramid pooling-fast is proposed to detect hematoma types, namely acute intraparenchymal, intraventricular, subdural, epidural, subarachnoid, and chronic subdural. Firstly, the computed tomography images are pre-processed using a window-based stacking approach wherein a three-channel image is generated by stacking grayscale images obtained with the aid of multiple window settings, i.e., brain, bone, and subdural. Secondly, a cascaded attention module is constructed in the neck of the YOLOv5s model to improve its detection performance by placing the convolution block attention module in serial with the efficient channel attention module. The cascaded attention module enriches the feature representation of various hematoma types in complex backgrounds especially when they are small and inconspicuous. The spatial pyramid pooling is replaced by a spatial pyramid pooling-fast to reduce the computational parameters and accelerate the feature fusion ability. The proposed deep learning model is trained, validated, and tested with 15,921 images from the brain haemorrhage extended dataset and it achieved overall precision, recall, F1-score, and mean average precision at 0.5, and mean average precision at 0.5:0.95 of 0.935, 0.908, 0.921, 0.943 and 0.65 respectively. The experimental results show that in comparison to the original YOLOv5s model and state-of-the-art methods, the model was able to localize and classify the acute or chronic instances of five hematoma subtypes in an individual image with improved precision and recall values. Hence the proposed system can be used in hospitals for the early and accurate detection of hematoma.

**INDEX TERMS** Cascaded attention module, medical image processing, intracranial hematoma, TBI, YOLO.

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Bouchir<sup>1</sup>.

## I. INTRODUCTION

Traumatic Brain Injury (TBI) is a complex medical disorder that can cause temporary or permanent brain damage and

dysfunction due to external force [1]. TBI is one of the leading global health issues with 70 million people suffering from it annually, resulting in a total of 8.1 million years of life with disability [2]. The age-standardized frequency of TBI increased by 8.6% from 1990 to 2016 and the major reasons can be attributed to road accidents and fall [2]. Due to the impact of the brain injury, the cerebral blood vessels burst, and blood accumulates in the intracranial compartments, leading to Intracranial Hematoma (ICH). ICH is a devastating health problem with a long-term functional independence of 12-39% and a 30-day mortality rate of 40% [3]. Early detection and management of ICH can result in improved survival and recovery rates of the patients. Hence, prompt and accurate detection of ICH by clinicians is vital for better decision-making and patient outcomes. The hematoma can be classified into five major types which include Subarachnoid Hematoma (SAH), Subdural Hematoma (SDH), Epidural Hematoma (EDH), Intraparenchymal Hematoma (IPH) and Intraventricular Hematoma (IVH). The hematoma can appear either in an acute or chronic stage. The acute hematoma manifests rapidly within the first few hours of injury whereas the chronic hematoma evolves over weeks after the injury. SDH is the most common hematoma in pediatric and elderly patients with different phases of manifestations [4], [5].

Non-contrast Computed Tomography (CT) is the gold standard imaging modality used to assess hematoma after the incidence of acute TBI. CT imaging is preferred to other imaging modalities for the initial diagnosis of ICH as it is fast, economical, widely available, and highly sensitive to blood and bone tissues [6], [7], [8]. The hematoma can be observed as hyperdense regions in the CT images with highly irregular shapes and sizes, especially during the stage of hematoma expansion. Further, the hematoma regions can be observed in a acute or chronic phase. The CT images are inspected by skilled radiologists to detect and classify the hematoma regions. However, due to the intrinsic limitations in the grayscale CT images and extreme variabilities with respect to shape, size, and location, manual hematoma detection is arduous and requires high expertise [9], [10], [11], [12]. Therefore, Computer-Aided Diagnostic (CAD) techniques that facilitates fast and accurate detection of hematoma is essential to provide improved quality care to the patients.

## A. LITERATURE REVIEW

Numerous CAD systems have been developed in the last two decades for the segmentation, classification, and detection of ICH. The CAD systems are aimed at providing rapid, low-cost, and reliable detection of hematoma by capturing varying levels of powerful discerning features in the CT images. Some of the conventional image processing techniques were exclusively based on either segmentation or classification techniques. For example, Shahangian and Pourghassem [13] applied a modified distance regularized level set evolution method to initially obtain the hematoma regions and then applied a hierarchical classifier to segment and categorize

ICH. With the optimal set of shape and texture features, the classifier achieved an accuracy of 94.13% for the characterization of four hematoma types. Kumar et al. [14] integrated fuzzy c-means clustering, entropy-based thresholding, and distance-regularized level set methods for the segmentation of hematoma. The technique attained an accuracy of 99.87% for a dataset of 35 CT images. Alawad et al. [15] evaluated the performance of a hematoma classification framework using Otsu's thresholding, region growing, genetic algorithm, and various stacking models. The framework obtained an accuracy of 99.5% for four-class labelling using a Support Vector Machine (SVM) classifier.

Various deep-learning techniques were successfully applied in recent years for hematoma segmentation/classification/segmentation and classification. The segmentation techniques enables clinicians to assess the hematoma volume more accurately and guide towards surgical planning and better decision-making. Inkaew et al. [16] proposed a modified DeepMedic architecture which includes multiple parallel pathways for accurate segmentation of hematoma regions. The pre-processed CT images using brain window and subdural window are used to develop the model, and post-processing is performed for finer segmentation results. The method obtained a median Dice coefficient above 0.37 for hematoma segmentation. However, the model addressed only three types of hematoma namely SDH, EDH, and IPH. Phaphuangwittayakul et al. [17] applied an interated approach using EfficientNetB2, optimal DeepMedic model, and assessment algorithm for segmentation, classification, and quantification of hematoma volume. The DICOM images are scaled to three different windows and the three scaled images are combined to obtain the three-channel image. The model achieved a sensitivity, specificity, and accuracy of 96.01%, 97.55%, and 96.54% respectively. However, the model can perform only three-class detection and images with multiple hematoma patterns of various classes are not considered. Ganeshkumar et al. [18] presented an ICH diagnosis model based on Residual Neural Network (ResNet) for detection and Adversarial Network (SegAN) for segmentation. With the aid of a novel augmentation method named Cycle Generate Adversarial Network (CycleGAN), the method obtained an average sensitivity of 0.80 and specificity of 0.99 for 82 CT scans from the PhysioNet database. However, CT images with various ICH subtypes and mixed cases were not taken into account. Ma et al. [19] focused on evaluating the performance of a modified U-Net model with an attention mechanism for hematoma segmentation. To capture the tiny hematoma regions, Residual Hybrid Atrous Convolution Strategy (RHAC) modules are included in every stage of the encoder. Moreover, by introducing feature max-pooling with varied sizes and an attention mechanism to capture the semantic features channel-wise and spatial-wise, the model achieved sensitivity and specificity of 75% and 99% respectively. However, the hematoma subtype is not taken into account.

Besides the various classification and segmentation techniques, a few object detection techniques were reported recently for hematoma diagnosis. Ertugrul et al. [20] proposed a YOLOv4 model for the detection of six hematoma classes using the BHX dataset. The model was trained separately with two different datasets - one class label per image and multiple class labels per image. The model obtained precision, recall, and mean Average Precision at 0.5 (mAP@0.5) of 93.8%, 91.8%, and 90.6% respectively for a set of images with one label. But the precision, recall, and mAP@0.5 values were reduced to 92%, 81% and 79.6% during training with multiple labels per image. Kothala et al. [21] proposed a light-weight YOLOv5x-Ghost Convolution Block(GCB) for hematoma detection using the BHX dataset. To reduce the computational parameters and improve the detection performance, all of the CSP bottleneck structures in YOLOv5x were replaced by a ghost bottleneck that utilizes ghost convolution. In order to further enhance the efficiency and speed of ICH detection, the Spatial Pyramid Pooling-Fast(SPPF) module is utilized in the proposed scheme. The model was able to detect the co-existence of multiple hematoma types in a CT image with precision, recall, and mAP@0.5 of 92.1%, 88.9%, and 93.1% respectively. However, the inclusion of a ghost module in the YOLOv5x has brought down the performance of the model. The precision and recall rates of the model can be further improved for the detection of mixed images with single or multiple patterns of hematoma types.

## B. LITERATURE GAPS

Based on the above literature review, the literature gaps are as follows:

- a. For Traditional Techniques:
  - They uses pre-defined values, and rules, requiring adjustment of various control parameters, making them highly unreliable for diverse, real-time clinical cases.
  - They uses limited dataset size and detect few classes of hematoma.
  - They require complicated steps such as spatial image registration, skull removal, and voxel-level details for hematoma detection.
  - Some are computationally expensive, sensitive to initialization, and require a large number of iterations to converge.
  - They better detect medium and large hematoma lesions as compared to small, subtle, and isodense hematoma regions.
- b. For Deep Learning Techniques:
  - Most focused on either classification or segmentation of a few hematoma subtypes. Few works perform both localization and classification for five hematoma subtypes.
  - The size of data for some ICH subtypes is limited, thereby reducing the generalizability of the application.
- Few works have been published to detect acute and chronic hematoma lesions which requires immediate surgical intervention.
- Most of the studies are only capable of identifying or classifying a single type of hematoma in a CT image, even when multiple hematoma subtypes are present in various sizes and shapes.

## C. MOTIVATION

The motivation of the proposed study is outlined as follows:

- 1) Due to the presence of noise, artefacts, asymmetrical limits, and similar pixel intensity regions in the CT images, hematoma detection is a highly challenging and time-consuming process even for skilled radiologists [9], [10]. Further, manual examination and estimation is a highly operator-dependent task and is subjected to inter- and intra-observer variabilities. Additionally, such a labour intensive process may introduce faults and deferments [22], particularly in a large clinical environment. Moreover, various instances of misdiagnosis and clinical consequences were reported by the junior residents during the absence of skilled experts, particularly during odd hours [11], [12]. Therefore, CAD techniques that enable clinicians to perform prompt and accurate identification and assessment of hematoma are crucial, especially in remote geographic locations where expertise is scarce.
- 2) Most of the traditional CAD techniques suffer from numerous disadvantages like hardcoded logic, manual intervention, and reduced performance which makes them highly unreliable to handle diverse and heterogeneous cases in a real-time clinical environment [13], [14].
- 3) The standard deep learning-based two-stage object detection models (Region Based CNN (R-CNN), Fast R-CNN and Faster R-CNN) have various limitations.
  - a. The object detection model based on R-CNN uses a large number of region proposals for accurate detection thereby increasing the computational and training times [23], [24].
  - b. Although Fast R-CNN resulted in improved performance with better speed, its real-time application is limited due to poor region selection techniques [23], [25].
  - c. Despite the improved detection rates achieved by the Faster R-CNN, the requirements of high computational power and larger datasets make it less feasible for various applications [23], [26].
- 4) The models which belong to the single-stage YOLO series can be deployed easily with fewer computational parameters which makes them more suitable to meet real-time requirements [23], [27].
  - a. The Darknet-based YOLOv3 [28] was able to detect objects of various scales with improved detection accuracy especially for small objects.

- b. The YOLOv4 model was designed with CSPDarknet-53, Path Aggregation Network (PANet) and Spatial Pyramid Pooling (SPP) to obtain greater detection speed with better detection accuracy [29].
  - c. As compared to previous YOLO models, the inclusion of Feature Pyramid Network (FPN) and PANet structures along with CSPDarknet-53 enables YOLOv5 [30] model to achieve higher detection performance and speed.
  - d. The YOLOv5 model is implemented using PyTorch instead of Darknet and hence the model size is very small as compared to YOLOv4 [23], [30]. Due to the relatively smaller structure, improved detection accuracy, and speed, YOLOv5 is chosen as the basic framework for the hematoma detection model.
- 5) The hematoma regions can appear as a mixture of single or different patterns with highly varying shapes, sizes, and locations [31], [32]. Hence the automated detection process is a unique challenge especially when the hematoma is subtle and small [31], [32], [33]. Therefore, the major motivation of the research work is to localize and classify a single or combination of different hematoma patterns with irregular structures in the CT images. However, increasing the complexity of the network will make the training process more difficult and time-consuming. A deep learning network that is too simple will fail to capture the critical features of hematoma thereby reducing the detection accuracy. To solve the above problems and optimize the detection performance while reducing the computational complexity and detection time, a YOLOv5s- Cascaded Attention Module (CAM) is proposed.

#### D. CONTRIBUTIONS

The main contribution of the present work is presented as follows:

- A window-based stacking approach in the pre-processing stage incorporates multiple CT windows to improve detection of subtle hematoma regions and overall accuracy.
- The SPP in YOLOv5s is replaced by SPPF to enhance the speed and receptive field of the network for multiscale detection.
- A Cascaded Attention Module which combines the Efficient Channel Attention Module (ECA) and Convolution Block Attention Module (CBAM) is included in each of the feature fusion layers of the YOLOv5s model to increase the attention to highly significant channels and the corresponding spatial features. This helps to reduce the disturbance caused by overlapping regions and enhances the feature fusion ability to detect small and subtle hematomas.
- Based on the above improvements, a novel YOLOv5s-CAM is proposed to a) detect single or multiple hematoma patterns of the same or different types

- b) distinguish acute and chronic instances of SDH, which outperforms the existing state-of-the-art models.
- Model performance is evaluated using the shuffle split cross-validation technique, ensuring the ability to generalize in real-world scenarios.

The paper is structured as follows: Section II presents the detection principles of YOLOv5 and the enhancements that are added to improve the performance for hematoma diagnosis. Section III describes the details of the research experiment for hematoma detection. Section IV showcases the results obtained by the proposed model. Section V provides a comparative discussion of the results, and Section VI concludes the paper.

## II. YOLOv5 DETECTION PRINCIPLES AND IMPROVEMENT

### A. YOLOv5

YOLOv5 is the latest in the series of YOLO models which possess rapid inference time, high accuracy, and fewer storage requirements as compared to YOLOv4 [30]. The YOLOv5 can be implemented in five ways: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x [30]. These models vary based on the CSP structures, number of convolution kernels, and residual blocks in the architecture. The architecture includes four major units, namely the input, backbone, neck, and head [30]. The input unit is responsible for enhancing the input images and optimizing the anchor boxes. Mosaic data augmentation is applied to input images where four random images are cropped and combined to create new images and then they are scaled to a uniform size [30]. The mosaic augmentation technique enriches the input feature maps which will make the model robust for hematoma detection. The initial size of the anchor boxes is optimized for the training data by comparing the size of the predicted and the actual bounding boxes [34].

The backbone acts as a feature extraction unit that collects meaningful information from various feature maps with fewer computations and redundancy. The backbone consists of Focus, Convolution, Batch Normalization and SiLU Activation function (CBS), Cross-Stage Partial (CSP) networks, and SPP [34]. The focus layer slices the input image into four  $3 \times 128 \times 128$  images which are further concatenated to generate  $12 \times 128 \times 128$  slices [30], [34]. The CSP1 structure concatenates the feature maps from the multiple residual blocks and three convolution layers to perform improved feature extraction. The CSP1 network reduces the repetitive gradient information to solve the problem of gradient vanishing, especially for deeper networks [34]. The CSP structure enhances the inference speed and accuracy while decreasing the computation parameters and model size. The SPP layer performs max-pooling operations of four varied sizes on the input feature maps and concatenates the outputs from all four units to enhance the receptive field of the network [23].

The neck is the feature aggregation module which combines the features from various backbone and prediction layers. It consists of FPN, PANet, and CSP2 networks [30].



The FPN transfers the semantic information from top to bottom layers whereas PANet propagates the localization details from bottom to top layers [35]. The PANet and FPN structures enable multiscale detection of the target regions by combining strong semantic and localization features from top-down and bottom-up layers [30]. The CSP2 network further enhances the feature fusion capability by cascading  $X$  number of CBS operations [23].

The YOLO head makes the final detections by applying  $1 \times 1$  convolution to the refined feature maps from the neck [34]. It consists of three convolution layers to detect target objects of three different scales. The input feature map is divided into  $S \times S$  grids and generates  $B$  bounding boxes for each grid cell [36]. Hence the detection result consists of the coordinates of the bounding box, the predicted class of the target object, and the confidence score of the predicted class. To remove multiple bounding boxes predicted around the same object using the three detection layers, Non-Maximal Suppression (NMS) is applied with a confidence threshold of 0.25 and IoU threshold of 0.5 [37]. The loss function of the YOLOv5 is the summation of the classification loss function, confidence loss function, and regression loss function [38]. Since the research study aims to develop a rapid and accurate CNN model with limited resources for hematoma detection, YOLOv5s is chosen for further implementation and experimentation. The hematoma regions in the CT slices can be highly irregular with varied shapes and sizes. Further, the hematoma lesions look very similar in texture and shape, and hence detection of a mixture of single or multiple patterns of ICH subtypes, particularly in an individual CT image, is a rigorous problem. Thus, in order to improve the performance of the YOLOv5 network for better patient outcomes, additional optimization is required. The architecture of the optimized YOLOv5s deep learning model with the input of dimensions  $256 \times 256$ , backbone, neck, head, and the novel cascaded attention module (YOLOv5s-CAM) for hematoma detection is represented in Figure 1.

## B. IMPROVED YOLOv5s-CAM FOR ICH DETECTION

Due to the irregular structures and overlapping pixel intensity values of hematoma regions with the other brain tissues, detection of hematoma is very challenging even for an expert radiologist. Therefore, the YOLOv5s model needs to be improved to focus on essential features for accurate and rapid detection of diverse hematoma regions in the CT images especially when subtle multiple hematoma patterns co-exist. The unique features of the proposed YOLOv5s-CAM model are as follows: (a) A window-based stacking approach is employed as a pre-processing technique to enhance the detection of subtle changes in the input CT images (b) A cascaded attention module is incorporated in the neck of YOLOv5s to improve the model's ability to focus on key features, minimize the interference due to background objects and enrich the feature fusion ability to detect small and subtle hematoma by focusing on highly significant information

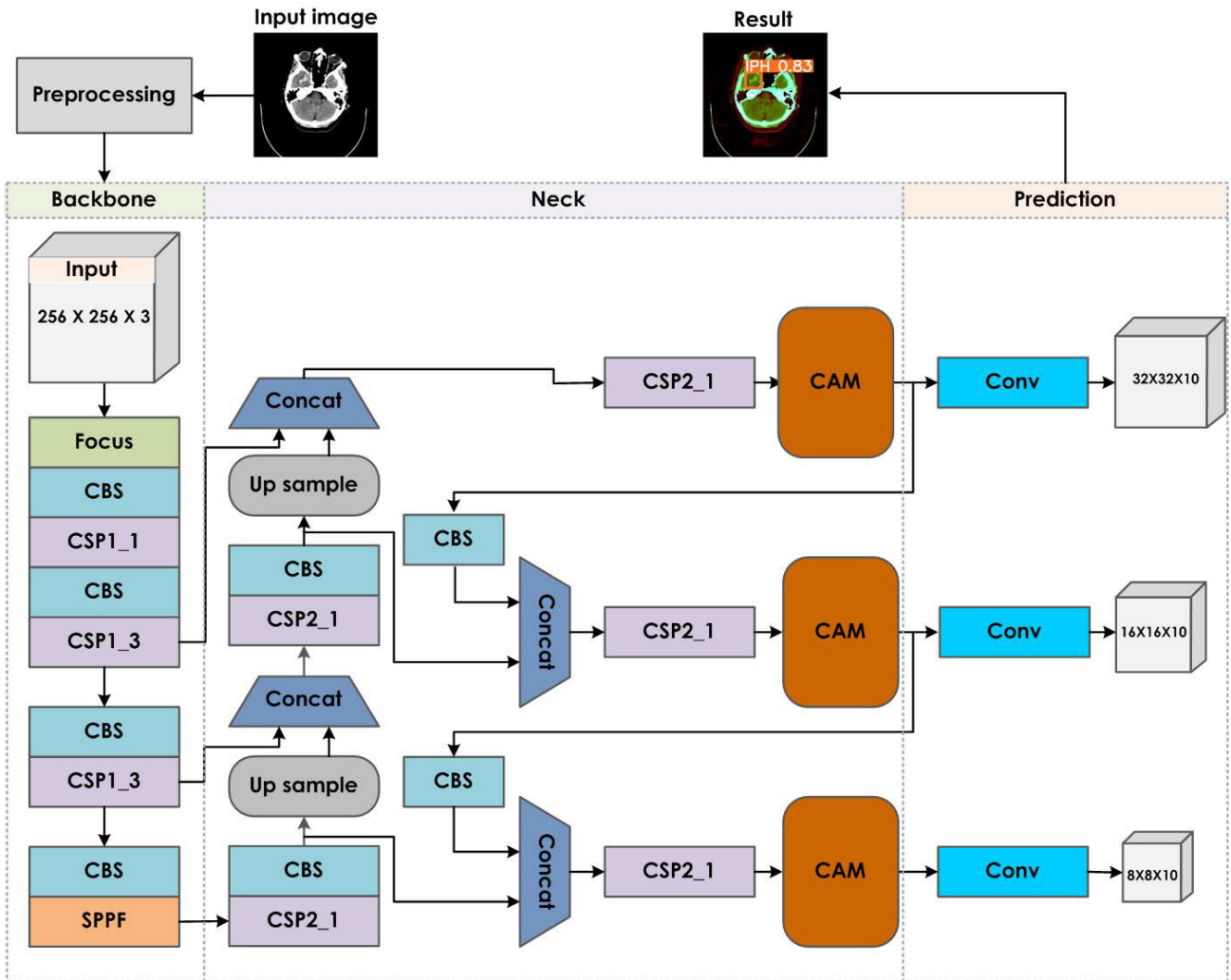
from vast amount of data. (c) The SPP module in YOLOv5s is replaced by SPPF to enhance the efficiency and speed of the model.

### 1) STACKING OF MULTIPLE WINDOWS FOR PRE-PROCESSING

Pre-processing is applied to the CT images in the digital imaging and communications in medicine (DICOM) format to discard the irrelevant details and extract relevant information. Pre-processing is a crucial step to enhance various attributes of the input images which further aids in more accurate and reliable predictions. In CT images, Hounsfield Units (HU) are used to represent the amount of X-ray absorption by the brain tissues [39], [40]. As the intensity range of HU values cannot be completely captured by the digital monitor, a windowing technique is applied by the radiologists to highlight the regions of interest in the CT images. Windowing facilitates the mapping of a selected interval of HU values into a grayscale range of 0 to 255 [41]. Two parameters that play a major role in the windowing process are window width and window level respectively. Window width represents the range of HU values that need to be converted to an entire grayscale limit of [0, 255] and window level indicates the middle value in the range of HU values. Window level ( $L$ ) allows highlighting the region of interest in the brain whereas window width ( $W$ ) is used to adjust the contrast of the selected region of interest [39], [42]. The HU values in the range of  $[L - 0.5 \times W, L + 0.5 \times W]$  are converted to [0, 255] and the values below  $L - 0.5 \times W$  and above  $L + 0.5 \times W$  are assigned black and white respectively [33]. However one window setting may not be adequate to detect the diverse abnormalities in the brain, and hence radiologists use multiple window settings for better detection and interpretation of CT images [41], [42]. Therefore, we follow the interpretation strategy of the radiologists as illustrated in Figure 2. The DICOM image is scaled to three different windows - bone window, brain window, and subdural windows with ( $L, W$ ) values as (500, 3000), (40, 80), and (175, 50) respectively. Then the three windowed images are assembled to acquire a three-channel RGB image [40] of size  $256 \times 256 \times 3$ . The creation of this RGB image maximizes the information provided to the network and ensures that the model can detect hematoma subtypes accurately. The obtained three-channel image is further used for training the improved YOLOv5 model.

### 2) SPATIAL PYRAMID POOLING-FAST

The purpose of SPPF is to enable multiscale fusion by generating and integrating semantic features from the input feature maps. Max-pooling is performed multiple times to realize the feature fusion. The YOLOv5s model uses the SPP module in the backbone network for the generation and fusion of features to improve the detection performance as shown in Figure 3(a). SPP utilizes kernel sizes of  $1 \times 1, 5 \times 5, 9 \times 9$ ,



**FIGURE 1.** The architecture of the proposed YOLOv5s-CAM model. The input is a DICOM image of size  $256 \times 256$ . The CAM module is placed in each of the three feature fusion layers in the neck. The outputs from the head include the prediction of hematoma with bounding boxes, class labels, and confidence scores at three different scales.

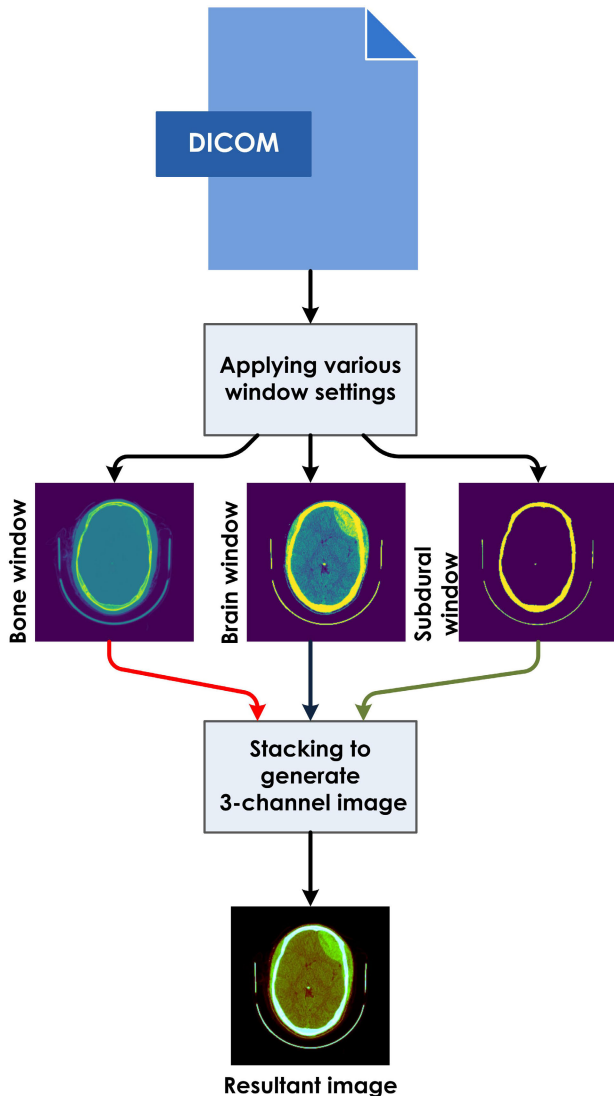
and  $13 \times 13$  to perform parallel max-pooling which requires resizing of the images multiple times to generate feature maps of fixed dimensions [23], [30], [35]. Further, the resizing leads to a greater loss of information, especially related to small target detection with overlapping of surrounding tissues. Hence the model may miss some of the detections resulting in reduced accuracy. Therefore, a SPPF structure which thrice performs pooling operations [35], [43] with a kernel size of  $8 \times 8$  is used in the proposed model instead of SPP. As shown in Figure 3(b), SPPF serially applies an  $8 \times 8$  max-pooling operation on the input data from the CBS structure. Hence the SPPF structure reduces computational parameters, improves the computational speed, and enhances the prediction accuracy for hematoma [35], [43].

### 3) CASCADED ATTENTION MODULE (CAM)

The attention mechanism is a strategy to maximize the performance of the deep learning model while minimizing

the computational complexity of processing the images [44]. Attention mechanisms are widely used in various deep learning applications due to their ability to provide better interpretations on complicated information by enabling more focus to the precise areas of the feature maps [23], [44].

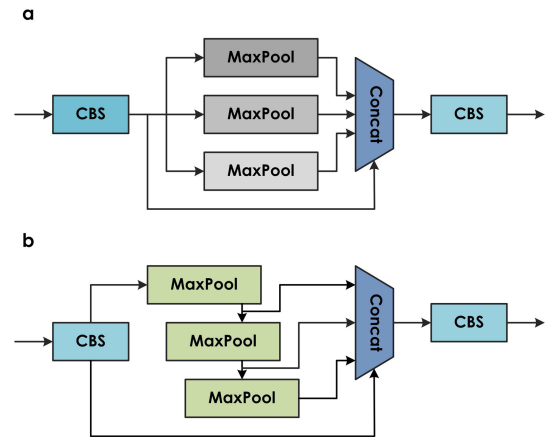
Hematoma detection in CT images is extremely difficult and complex even for expert clinicians because of the uneven structures and similar pixel intensities of hematoma regions with the surrounding regions. Thus, it is necessary to enhance the YOLOv5s model in order to concentrate on crucial characteristics for precise and quick identification of various hematoma locations in the CT scans. The inclusion of attention modules in the network enables the model to emphasize the key areas of the image by assigning more weights as compared to the other regions of the image [45]. Therefore, a cascaded attention mechanism that incorporates ECA and CBAM attention modules is proposed as depicted in Figure 4. The ECA attention module [46], [47] is placed



**FIGURE 2.** Outline of the windowing technique for pre-processing. Each DICOM image is scaled to brain window, bone window, and subdural window and further assembled to generate a 3-channel image.

first to select the key channels from the large volume of feature information without dimensionality reduction. The key channels are further amplified using the CBAM attention module [48]. The channel attention module refines the initial set of significant channels to obtain more profound and contributing channels. Then the spatial attention module as part of CBAM will capture spatial-wise information of the selected channels to obtain the final weighted feature map [48], [49]. The weighted feature map is then concatenated with the input feature maps to obtain the final set of significant feature maps for accurate multi-scale detection of hematoma in varied sizes and shapes.

The ECA attention mechanism [46] is included initially to capture the dependency among the various channels in the pre-processed CT images with reduced parameters, complexity and improved performance [34]. The channel



**FIGURE 3.** (a) Structure of spatial pyramid pooling (b) Structure of spatial pyramid pooling-fast.

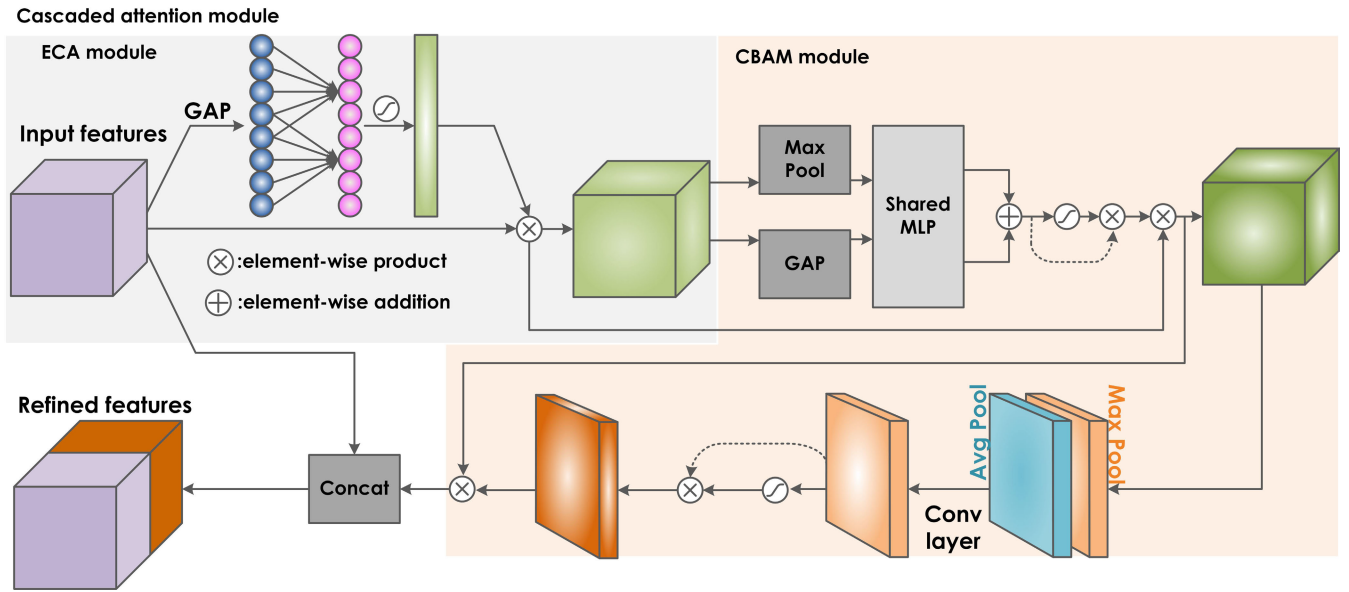
attention module uses a neural network to capture the cross-channel interactions which will reduce the dimensions of the model to reduce the complexity. However the dimensionality reduction reduces the ability to capture the cross-channel interactions effectively [46], [50]. Hence the ECA [46] module replaces the fully connected neural network with the fast 1D convolution of adaptive kernel size  $k$  to capture the weights after applying Global Average Pooling (GAP) to the input feature maps. The adaptive kernel size  $k$  enables one to decide the extent of cross-channel interactions based on the channel dimension  $C$  as given below [46], [47].

$$k = \phi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (1)$$

where  $\gamma = 2$ ,  $b = 1$  and  $\lfloor v \rfloor_{\text{odd}}$  denotes the nearest odd value to  $v$ . Next, the set of channel weights is produced by further applying the sigmoid function [51]. Thereafter the channel weights are multiplied with the input feature maps to obtain the refined set of feature maps  $M$  which will be provided as input to the next phase of the CAM structure.

In this phase, we have placed the CBAM attention module in series with the ECA module to further refine and strengthen the obtained channel feature maps along the channel and spatial dimensions simply and effectively. This serial arrangement of the CBAM module facilitates more attention to the significant channel maps with reduced parameters. Thus, a balance can be achieved by the model between the reduction of complexity, computation, and resources, and performance improvement. As shown in Figure 4, CBAM [48] consists of two units in succession – channel attention module and spatial attention module (SAM) [48]. The channel attention module executes average pooling and max pooling globally on the individual feature maps  $M$  of dimension  $C \times H \times W$  where  $C, H$ , and  $W$  indicate the number of channels, height, and width of the feature maps respectively [48], [49]. The two obtained vectors of size  $1 \times 1 \times C$  are passed to a two-layered neural network and the results are added together element-by-element, and the final





**FIGURE 4.** The structure of the proposed Cascaded Attention Module. The ECA attention mechanism is applied first and then followed by CBAM technique. The obtained weighted feature map is concatenated with the input feature map to generate significant set of features.

weighted feature map  $F_c$  of size,  $C \times 1 \times 1$  is obtained after applying the SiLU activation function [42]. Then the feature map  $F_c$  of size,  $C \times H \times W$  is produced by multiplying  $F_c$  with the input feature map  $M$ , which can be defined as follows [48], [49]:

$$Q_1 = F_c(M) \times M \tag{2}$$

$$F_c(M) = SiLU (W_2(W_1(M_c^{avg})) + W_2(W_1(M_c^{max}))) \tag{3}$$

Then  $Q_1$  is provided as input to the SAM module. SAM initially concatenates the feature maps obtained after applying global maximum pooling and global average pooling to  $Q_1$  separately, and then convolution  $f$  with kernel size  $7 \times 7$  is performed to adjust the number of channels. Finally, the SiLU activation function is used to obtain the feature map  $F_s$  of size  $1 \times H \times W$  [47], [48]. Then the final feature map  $Q_2$  of size  $C \times H \times W$  is generated by multiplying  $F_s$  with the feature map  $Q_1$  which can be stated as follows

$$Q_2 = F_s(Q_1) \times M \tag{4}$$

$$F_s(Q_1) = SiLU (f^{7 \times 7} [Q_1^{max}; Q_1^{avg}]) \tag{5}$$

The feature map  $Q_2$  is further concatenated with the input of the cascaded attention module to obtain the final set of refined feature maps which will be provided to the head for making the predictions. The head detects the hematoma regions and outputs three different feature maps of dimensions  $8 \times 8 \times 10$ ,  $16 \times 16 \times 10$ , and  $32 \times 32 \times 10$  respectively which will enable the detection of hematomas of large, medium, and small sizes. The detected hematoma regions are marked with bounding boxes along with the target class label and confidence score. To remove multiple bounding boxes predicted around the same object using the three detection layers, NMS is applied with a confidence threshold of 0.25 and IoU threshold of

0.5 [37]. Then, the bounding box with the highest confidence score is selected and the IoU of the remaining bounding boxes is computed with the selected box [38]. All bounding boxes with an IoU greater than the threshold value of 0.5 are removed. This process is repeated until all overlapping bounding boxes with a lesser confidence score are eliminated to obtain the final bounding box with highest confidence score around the object detected [37], [38].

### III. EXPERIMENT AND ANALYSIS

#### A. EXPERIMENTAL DATASET

The publicly available Brain Haemorrhage Extended (BHX) dataset [52] with bounding box annotations for six hematoma subtypes is used to perform the research study. BHX is an extended version of the CQ500 dataset which consists of 491 CT scans with 205 hematoma cases [53]. BHX includes 39,668 bounding box labels for six hematoma types which include five acute hematoma types namely IPH, IVH, SDH, SAH, and EDH, and one chronic SDH (CHR) [52]. The bounding box annotations are done for thick-slice and thin-slice image series either manually by three experienced radiologists or using interpolation. BHX is made available in three versions and the third version is the cleanest one which includes the bounding box annotations for the selected thin-slice image series. We have selected 15,921 images from the cleaner version of the dataset to generate optimal results. The number of bounding boxes for various hematoma subtypes is shown in Table 1. As observed in Table 1, the dataset includes 3672 bounding boxes of chronic SDH out of the total 27096 bounding boxes for all hematoma subtypes. Further the number of bounding boxes for acute SDH, SAH, IPH, IVH, and EDH subtypes in descending order are 7904, 7530, 5248, 2194, and 548 respectively. A set of sample

images from the BHX dataset indicating various hematoma subtypes is illustrated in Figure 5. From left to right in the top row of Figure 5, IPH represents a spherical mass within the brain parenchyma, SAH is observed as white patches or tracks at various fissures and cisterns in the cerebral sulci, and EDH is observed as a lentiform structure between the skull and dura mater. As shown in the bottom row from left to right in Figure 5, IVH bleeds within the ventricular system while SDH bleeds between the arachnoid and pia mater in crescent form [54], [55]. The acute and the chronic SDH is identified as hyperdense and hypodense regions in the CT images respectively [55]. However, the hematoma shapes vary from its initial shape to a highly irregular shape and structure during its course of expansion [56].

## B. EXPERIMENTAL SETUP

The proposed hematoma detection model was executed using the Google Colab Pro platform equipped with a Tesla T4 GPU, OpenCV 4.6.0, and PyTorch framework. The entire dataset was divided in the ratio of 6:2:2 randomly, resulting in 9552, 3184, and 3185 CT images for training, validation, and testing respectively. The YOLOv5s-CAM model was provided with  $256 \times 256$  images which will then pass through the backbone and neck layers that have learned features to locate hematoma regions with the appropriate labels. An image size of  $256 \times 256$  is selected to reduce the computational complexity and training time while preserving the computational performance of the model [57], [58]. The batch size during training of the model was set to 64 indicating the number of CT images processed by the model before updation of internal parameters [59]. The training process aids in fine-tuning the initial weights of the model while minimizing the loss function using the optimizer. The head of the YOLOv5 model performs hematoma prediction, including bounding boxes to locate the hematoma regions and the class label for the predicted subtype of hematoma along with the confidence score. The confidence threshold and IoU threshold were set to 0.25 and 0.5 respectively for hematoma prediction [59]. The initial number of epochs for training the proposed model was set to 50. Figure 6 shows the mAP@0.5 achieved by the proposed model when the epoch size is gradually increased during the training phase. It is evident from Figure 6 that the mAP@0.5 for the proposed model gradually increases with respect to the number of epochs and achieves the highest mAP of 94.3% when the epochs are set to 300. Hence the model was trained for 300 epochs.

The initial learning rate is set to 0.01 and the Stochastic Gradient Descent (SGD) along with reducing-learning-rate-on-plateau function [59], [60] are utilized to optimize the training process. The reducing-learning-rate-on-plateau function [59] reduces the initial learning rate if the training loss stops improving after some specific number of epochs. This will enable the model to move fast in the beginning and converge slowly in the later stages to obtain the final

learning rate of  $1.6 \times 10^{-4}$ . Table 2 summarizes the various hyperparameters used for training the model [30].

TABLE 1. Number of bounding boxes for various ICH subtypes.

ICH Subtype	No. of bounding boxes
acute IPH	5248
acute IVH	2194
acute EDH	548
acute SDH	7904
acute SAH	7530
chronic SDH	3672
Total	27096

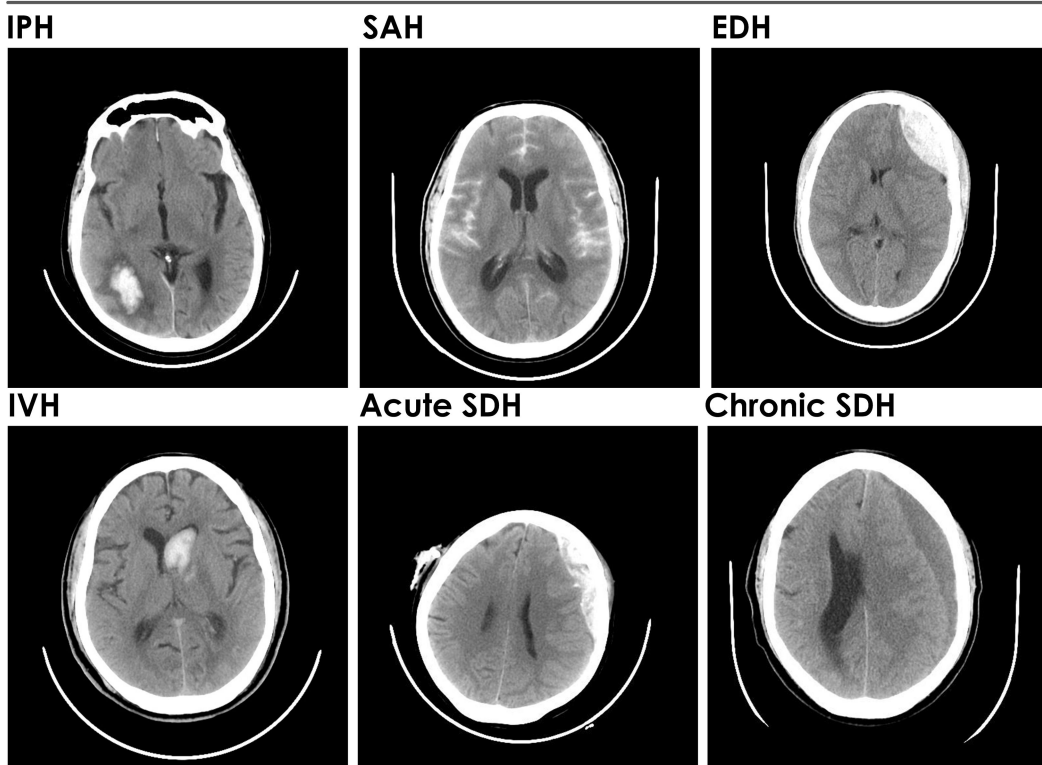
TABLE 2. Hyperparameters used for training the network.

Hyperparameters	Values
Training epochs	300
Batch Size	64
Momentum	0.937
Learning Rate	0.01
Decay	0.0005
Image Size	$256 \times 256$

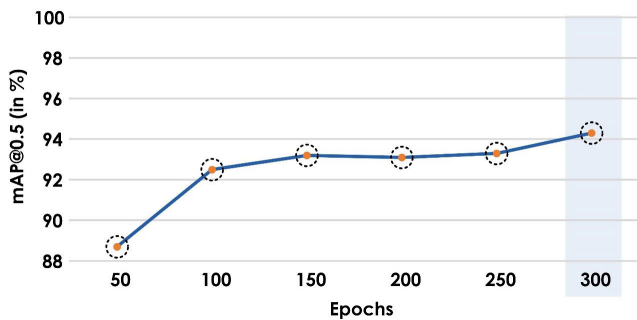
Figure 7 displays three distinct forms of loss curves for the training and validation set of ICH images based on the improved YOLOv5s-CAM model: box loss, object loss, and classification loss. Box loss indicates how effective the algorithm is in determining the centre of the object as well as the coverage of the predicted bounding boxes [23], [35]. Object loss denotes the capability of algorithm to detect an object in the selected area of interest. Classification loss represents the ability of the algorithm to predict the appropriate class of the target object. The  $x$  and  $y$  axes indicate the epochs and the corresponding loss values respectively [23], [35]. The total loss is the sum of box loss, classification loss, and object loss respectively [23], [35]. From the training and validation loss curves in Figure 7(a) and Figure 7(b), it can be observed that the loss values of the network are initially large, then experience a sharp decrease and gradually achieve stability after 50 epochs. Figure 7(c) shows the performance of the model by testing the validation set after each stage of training and the mAP@0.5 and mAP@0.5:0.95 values are recorded for every epoch. It can be observed from Figure 7(c) that the model achieved maximum mAP@0.5 and mAP@0.5:0.95 values of 94% and 64.93% once the network became stable. Therefore, we can further use the trained model which has achieved optimal mAP values for the validation set. Once the training and the validation steps were completed, the inferences are made using the best-trained weights on the test images for real-time hematoma detection.

## C. EVALUATION METRICS

The performance of the proposed hematoma detection model is evaluated using precision, recall, mAP@0.5,



**FIGURE 5.** Sample set of CT images from the BHX dataset. Top row from left to right - IPH as spherical mass within brain parenchyma, SAH as tracks in the cerebral sulci, EDH in lentiform shape between the skull and duramater. Bottom Row from left to right - IVH in the shape of ventricles within brain parenchyma, and acute SDH and chronic SDH as hyperdense and hypodense regions in crescent shape between arachnoid and pia mater.



**FIGURE 6.** Trend of mAP@0.5 during the training of the proposed model.

mAP@0.5:0.95 and F1-score which is computed based on the intersection over union (IoU) [23], [61]. IoU indicates the ratio of overlap between the predicted bounding box and the ground truth box. Then a detection can be considered as true positive (TP) if the predicted and the ground truth class labels match and the IoU threshold is above 50% [34], [62]. Conversely, if the class labels are matching but the threshold is less than 50%, then the detection is false positive (FP). But if the bounding box is not detected for the corresponding ground truth box, then the detection is false negative (FN) [49], [62].

**Precision (P)** can be defined as the ratio of correctly predicted detections to the total detections by the model [23], [62].

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

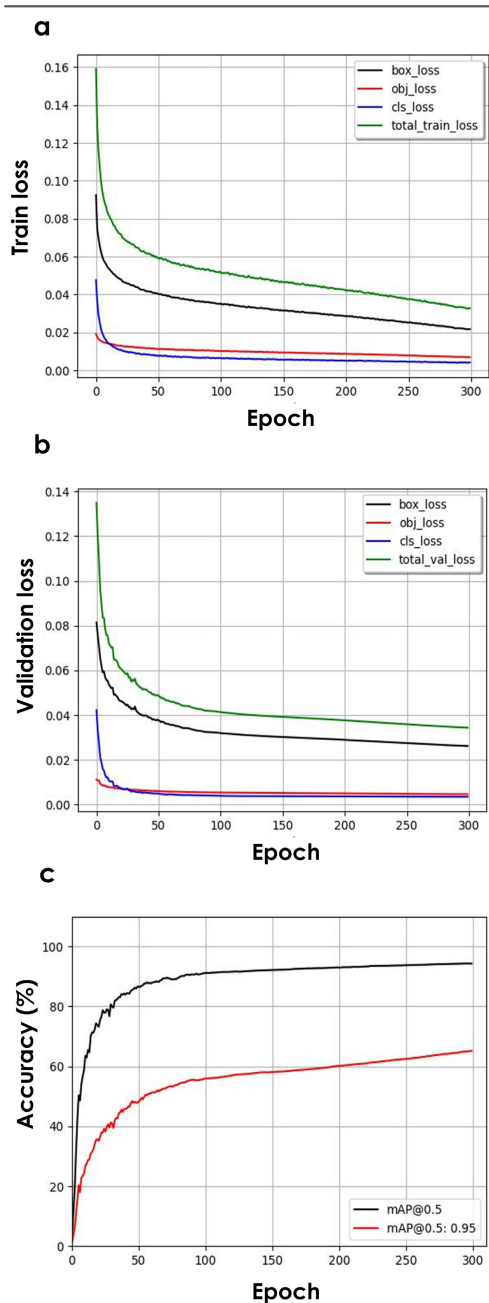
**Recall (R)** is the ratio of detections that are predicted correctly out of all the ground truth samples [21], [23], [62].

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

**F1-score** is a metric which uses precision and recall values to assess the performance of the model [21], [23].

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

The **Precision-recall curve** can be obtained by plotting the recall and precision values for each class in the x and y axes respectively [23], [62]. The area under the precision-recall curve is the average precision (AP) for that class. **mAP@0.5** is the average of the AP values of each class which is computed with an IoU threshold of 0.5 [20], [23], [62]. **mAP@0.5:0.95** denotes the mean of mAP@0.5 values obtained by using the IoU threshold range of 0.5 to 0.95 with an increment of 0.05 at each step.

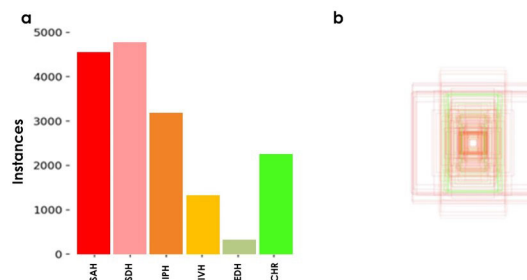


**FIGURE 7.** Comparison curves of loss functions (box loss, object loss, classification loss, and total loss) and mAP (a) Loss Curves for the training dataset. (b) Loss Curves for the validation dataset. (c) Test Curve for mAP@0.5 and mAP@0.5:0.95 with respect to epochs for the validation dataset.

#### IV. EXPERIMENTAL RESULTS

The proposed model is developed using 15,921 CT images from the BHX dataset in the DICOM format of size  $256 \times 256$ , resulting in 27,096 bounding boxes for six different subtypes of ICH. An analysis of various categories of hematoma and their bounding box information in the training dataset is illustrated in Figure 8. Figure 8(a) shows the class-wise distribution of hematoma subtypes with a relatively higher

representation of SDH instances and a lower representation of EDH instances based on the inherent limitation of EDH cases in the BHX dataset. Figure 8(b) depicts the actual shape and size of bounding boxes in the training set based on the ground truth information. It can be seen from Figure 8(b) that the target hematoma regions in the training dataset are mainly medium to small-sized objects which indicates the need for proper detection approaches.



**FIGURE 8.** Analysis of data in the training set (a) Plot showing the distribution of each category of ICH (b) Aspect ratios and dimensions of the ICH bounding boxes.

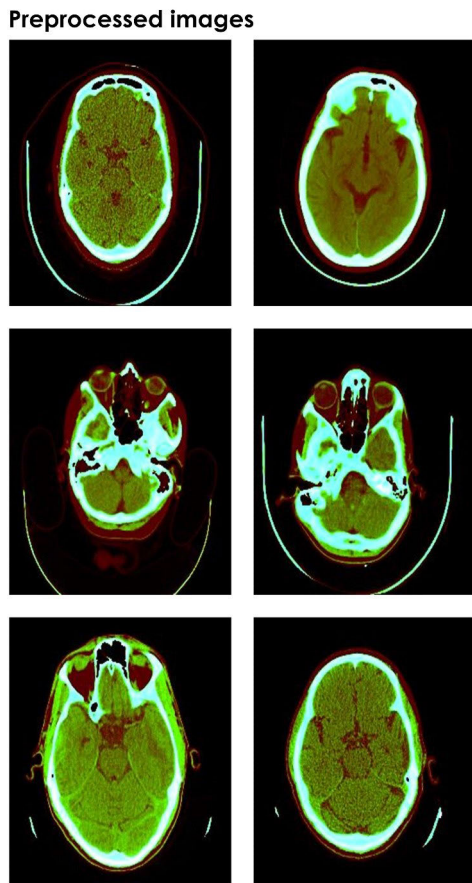
The CT images in the DICOM format are pre-processed initially to improve the overall performance of the system. The brain, bone, and subdural window settings are applied to the DICOM images to obtain the three grayscale images. Then the obtained grayscale images are stacked to generate the three-channel image of size  $256 \times 256 \times 3$ . The sample set of pre-processed images of six ICH subtypes is shown in Figure 9. It can be observed from Figure 9 that the pre-processed images provide better information about the subtle abnormalities which are unable to be distinguished from normal images by using a single window.

#### A. DETECTION RESULTS ON THE TEST DATA

The proposed model is trained to detect six different classes of hematoma. Figure 10 shows the results of the prediction by the proposed model on the test data. The predicted hematoma regions in the CT images are localized by bounding boxes of various colors indicating each ICH subtype. Further, the prediction results include the class label of the detected hematoma region with the confidence score. The confidence score indicates the probability of correctly predicting bounding boxes around objects, and how accurate the predicted bounding boxes are. Figure 10(a) and 10(e) show predicted instances of a mixture of single or multiple patterns of SAH, IVH, IPH and SAH respectively with improved confidence scores. Figure 10(b) and Figure 10(f) illustrate multiple chronic subdural types predicted by the model with confidence scores of 0.9, 0.9 and 0.94, and 0.95 respectively. Figure 10(c) and (d) depicts the predicted co-existence of IPH, SAH, and SDH, SAH with confidence scores of 0.9, 0.8, and 0.8, 0.8 respectively.

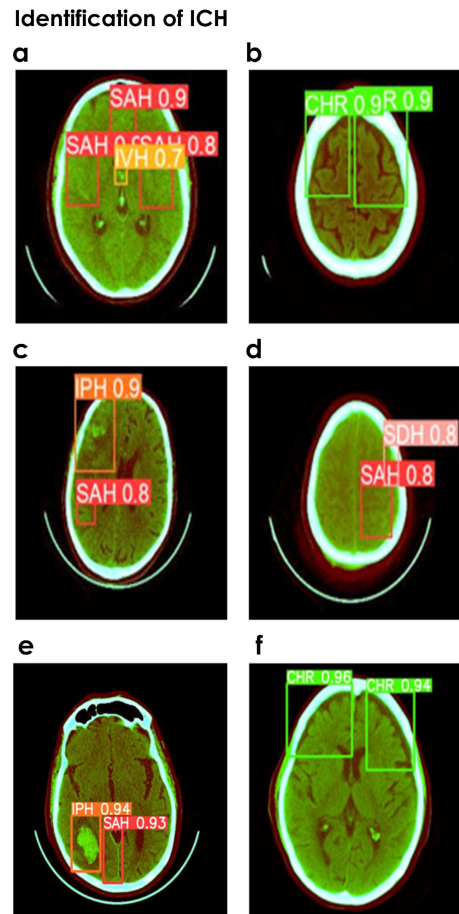
It can be observed from Figure 10 that the model was able to predict the various classes of hematoma.





**FIGURE 9.** Sample set of pre-processed images obtained by integrating the scaled DICOM images using brain window, bone window and subdural window.

It is evident from 10(a)-(e) that the model can detect the various hematoma subtypes present in a single CT image. Figure 10(f) shows that the model is capable of distinguishing the acute and chronic cases of SDH. Further, the model can detect the presence of multiple instances of a single hematoma as shown in Figures 10(a) and 10(b). The performance of the model to detect the six classes of hematoma is shown in Table 3. The precision, recall, and mAP@0.5 values range from 0 to 1. It is evident from the Table 3 that the proposed model was successful in detecting all six classes of hematoma with an overall precision, recall, and mAP value of 0.935, 0.908, and 0.943 respectively. Further, the model achieved the highest mAP of 0.994 for the detection of EDH with precision and recall values of 0.99 and 0.976 respectively. The model obtained the lowest mAP of 0.896 for the detection of IVH with precision and recall values of 0.888 and 0.846 respectively. As compared to EDH, the detection of IVH is complicated due to the presence of surrounding brain tissues which are of similar intensity, size, and shape. Moreover, the model was able to clearly distinguish the chronic and acute cases of SDH with precision and recall values of 0.93, 0.916, and 0.947, 0.943 respectively.



**FIGURE 10.** Sample set of detected images by the proposed model with predicted class label, confidence score, and bounding boxes around detected objects. (a) Multiple patterns of SAH and IVH with confidence score of 0.9, 0.8 and 0.7 (b) Multiple patterns of Chronic SDH with confidence score of 0.9 and 0.9 (c) IPH and SAH with confidence scores of 0.9 and 0.8 (d) SDH and SAH with confidence scores of 0.8 and 0.8 (e) IPH and SAH with confidence scores of 0.94, and 0.93 (f) Multiple patterns of Chronic SDH with confidence score of 0.95 and 0.94.

Despite the class imbalance shown in Table 1, the model was able to achieve remarkable and stable performance for all hematoma subtypes. The proposed model was able to detect the hematoma regions with an inference time of 1.2 ms per image.

The detection results for the six different classes are illustrated in Figure 11 using a confusion matrix. As shown in Figure 11, the accuracy of target detection rates of each type of hematoma is comparatively high while the miss rates are relatively low. The total number of TP for each hematoma subtype which can be observed on the diagonal of the confusion matrix in Figure 11 are as follows: 90% of all acute hematoma regions in the SAH subtype, 92% of all acute hematoma regions in the SDH subtype, 88% of all acute hematoma regions in the IVH subtype, 88% of all acute hematoma regions in the IPH subtype, 98% of all acute hematoma regions in the EDH subtype, and 86% of all



hematoma regions in the chronic SDH type were correctly detected and classified by the model. Further, 2% of all the SAH instances were misclassified as IPH, and 8% were not classified into any of the subtypes. 1% of all acute SDH instances were wrongly predicted as chronic SDH and 7% were not categorized into any hematoma subtypes. 2% of all IPH instances were incorrectly classified as SAH while 10% were not classified/identified by the model. 1% of all IVH instances were miscategorized as IPH and 10% were unrecognized. 2% of all EDH instances were unclassified by the model. 8% of all chronic SDH instances were mistakenly categorized as acute SDH while 6% were not categorized into any ICH subtype. Further, the background(BG) FP indicates the regions in CT images that does not belong to any of the six classes and detected as one of the ICH subtypes. The background FN indicates the regions not detected by the model and considered as non-hematoma regions.

Predicted	SAH	0.90		0.02				0.30
	SDH		0.92				0.08	0.31
	IPH	0.02		0.88	0.01			0.17
	IVH				0.88			0.09
	EDH					0.98		
	CHR		0.01				0.86	0.13
	BG	0.08	0.07	0.10	0.10	0.02	0.06	
			SAH	SDH	IPH	IVH	EDH	CHR
		True						

FIGURE 11. Performance evaluation of the proposed model using confusion matrix.

TABLE 3. Class-wise performance of the YOLOv5 model during detection.

ICH Subtype	Precision	Recall	mAP@0.5
SAH	0.908	0.891	0.923
SDH	0.93	0.916	0.952
IPH	0.944	0.878	0.927
IVH	0.888	0.846	0.896
EDH	0.99	0.976	0.994
CHR	0.947	0.943	0.966
Overall	0.935	0.908	0.943

Figure 12 illustrates the various curves which are obtained based on the confidence scores, precision, and recall values. Figure 12(a) and (b) shows that the proposed model was able to detect the six classes of hematoma with higher precision values while maintaining better recall values. Figure 12(c) clearly illustrates that the area under the curves for each of

the classes is greater, and EDH has the greatest area under the curve resulting in a mAP of 0.994. Further, the model was able to make significant predictions with an overall mAP of 0.943.

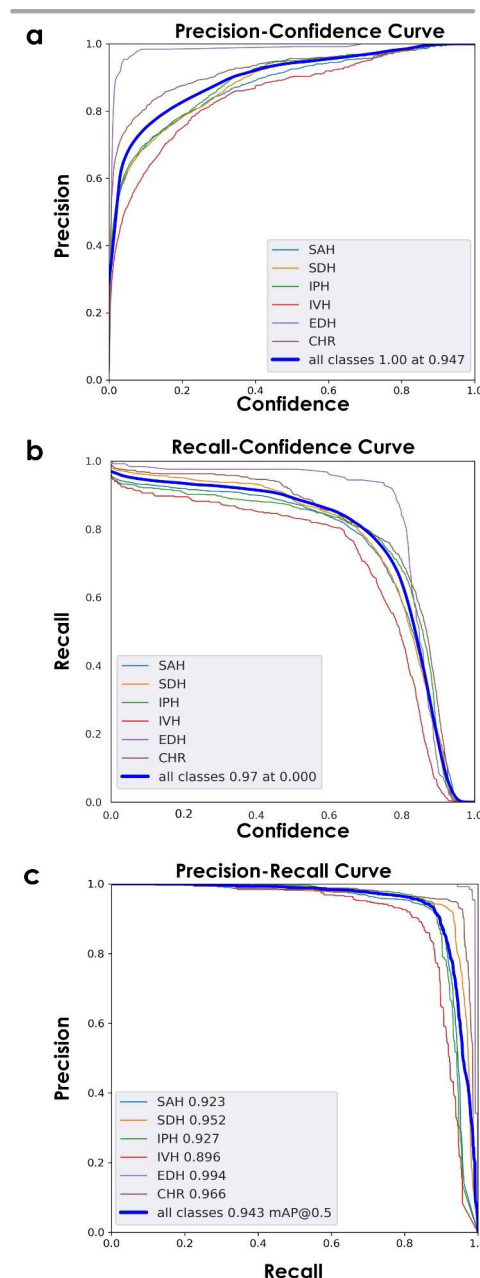


FIGURE 12. Curves that are plotted based on the confidence score and performance indicators. (a) Precision- confidence curve, (b) Recall- confidence curve, and (c) Precision-recall curve.

## V. DISCUSSION

The research study offers a robust deep-learning model that can rapidly and effectively localize and categorize six hematoma subtypes present in a single CT image or multiple CT images with precision, recall, and mAP values of 0.935, 0.908, and 0.943 respectively. Additionally, the proposed

model can discern the acute and chronic instances of subdural hematoma, which is one of the most frequent sequelae of TBI that requires immediate surgical intervention. The proposed model was developed using a comparatively larger dataset and was able to detect hematoma subtypes using fewer computational resources.

#### A. COMPARISON OF SINGLE WINDOW AND STACKED WINDOW APPROACHES

The proposed model uses a stacked window approach for the accurate detection of hematoma regions of various shapes and sizes. Initially, the single window setting was used to detect the hematoma regions in the CT images. Three different window settings, namely: brain window, bone window, and subdural window, were applied to the input CT images. To evaluate the impact of pre-processing, the YOLOv5s-CAM model is separately trained using single and combined window settings. The performance of the model is evaluated using the single window setting, by training and testing it with the three sets of windowed images separately. The three windowed images from each CT image are combined together to form a 3-channel RGB image (Figure 2) that is further utilized to assess the performance of the proposed hematoma detection model. The experimental results are shown in Table 4. The proposed model with the default brain window had the best mAP@0.5 of 0.933 as compared to the bone window (mAP@0.5 of 0.909) and subdural window (mAP@0.5 of 0.903) settings. However, as shown in Table 4, the mAP was further increased to 0.943 when the three different windowed images are combined. Therefore, we can conclude that the stacked window approach which mimics the radiologist mode of analyzing brain CT images, can be useful for efficient and accurate detection of subtle and inconspicuous hematoma regions.

**TABLE 4.** Performance of the YOLOv5s-CAM model using single window and stacked window approaches.

Sl.No	Model	mAP@0.5
1	YOLOv5s-CAM with brain window	0.933
2	YOLOv5s-CAM with bone window	0.909
3	YOLOv5s-CAM with subdural window	0.903
4	YOLOv5s-CAM with stacked windows	0.943

#### B. ABLATION STUDY

Various improvements were incorporated into the basic YOLOv5s model, and to study the effect of these improvements on the performance of the hematoma detection model, ablation studies were conducted. Table 5 summarizes the results of the ablation experiments on the YOLOv5 model using the test data.

The baseline version of the YOLOv5s model with SPP is compared initially with SPPF to identify the model that can generate a powerful set of refined features for hematoma

detection. However, the inclusion of SPPF in the network structure improves the mAP@0.5 and F1-score to 0.20% and 0.50% respectively as compared to SPP. Further experiments were carried out to enhance the model by focusing more on the target areas and suppressing irrelevant details with the aid of attention modules. The ECA attention module was incorporated in the neck of the YOLOv5s model, and it can be observed that the performance of the network with ECA and SPPF is higher than the ECA and SPP network, by improving precision, recall, mAP@0.5, and F1-score by 0.1%, 0.1%, 0.2%, and 0.10% respectively. This experiment indicates the effectiveness of the inclusion of ECA and SPPF modules in the proposed network. Further, to enrich the feature fusion capability, experiments were performed to analyze the performance of the YOLOv5s with the ECA model along with SAM and CBAM attention modules. The comparison results of these experiments were shown in (5), (6), (7), and (12) which shows that placement of the CBAM module as compared to SAM in series with ECA module achieved an improvement of precision and mAP@0.5 by 0.6%, and 0.6% respectively. Hence the CBAM module is also included along with the ECA module for strengthening the feature aggregation capability, which resulted in an improved mAP of 94.3%, thereby enhancing the performance of the YOLOv5s with SPPF network to 0.4%. Experiments (8-11) were carried out to further improve the model with the introduction of the microscale head [47] and the Double ECA attention module [63]. However, it can be concluded that the YOLOv5s model with the combination of SPPF, ECA, and CBAM can perform optimized hematoma detection as compared to experiments (9) and (11) with an increase in mAP of 0.2% and 0.3% respectively.

To further show the statistical significance of the proposed YOLOv5s-CAM model among the other models in the ablation study, One-Way Analysis of Variance (ANOVA) [64] is utilized. Firstly, the top three models such as YOLOv5s with SPP, YOLOv5s with ECA, CBAM and SPP, and YOLOv5s with ECA, CBAM, and SPPF are initially selected based on their performance (please refer to Table 5). Secondly, these three models are trained and tested five times each, and then mean values of obtained performance measures namely precision, recall, F1-score, and mAP@0.5 are calculated. Finally, One-Way ANOVA is applied to rank the mean performance measures of the selected three models. The proposed YOLOv5s-CAM model achieved a higher ranking than the other two, which indicates the statistical superiority of the model.

#### C. PERFORMANCE COMPARISON WITH CONTEMPORARY MODELS IN THE YOLO SERIES

For further quantitative analysis and performance assessment of hematoma detection, the proposed YOLOv5s-CAM model is compared with the YOLOv5s, YOLOv6s, and YOLOv7 algorithms, and the results are shown in Table 6. It can be

**TABLE 5. Results of the ablation experiments.**

Sl.No	Model	Precision	Recall	mAP@0.5	F1-Score
1	YOLOv5s with SPP	92.7	90.3	93.7	91.5
2	YOLOv5s with SPPF	93.1	90.9	93.9	92.0
3	YOLOv5s with ECA and SPP	92.4	90.9	93.7	91.6
4	YOLOv5s with ECA and SPPF	92.5	91.0	93.9	91.7
5	YOLOv5s with ECA, SAM, and SPP	92.7	91.0	93.8	91.8
6	YOLOv5s with ECA, SAM, and SPPF	92.9	91.1	94.0	92.0
7	YOLOv5s with ECA, CBAM, and SPP	93.4	90.7	94.2	92.0
8	YOLOv5s with Double ECA, CBAM, and SPP	92.3	90.7	93.8	91.5
9	YOLOv5s with Double ECA, CBAM, and SPPF	93.5	91.4	94.1	92.4
10	YOLOv5s with Microscale head, Double ECA, CBAM, and SPP	91.3	91.0	93.9	91.2
11	YOLOv5s with Microscale head, Double ECA, CBAM, and SPPF	92.0	90.6	94.0	91.1
12	<b>Proposed</b>	<b>93.5</b>	<b>90.8</b>	<b>94.3</b>	<b>92.1</b>

**TABLE 6. Performance comparison with contemporary models in the YOLO series.**

Sl.No	Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	F1-Score
1	YOLOv5s	92.7	90.3	93.7	64.5	91.5
2	YOLOv6s	92.6	83.0	89.1	61.1	87.6
3	YOLOv7s	48.1	40.2	41	18.5	43.8
4	<b>Proposed</b>	<b>93.5</b>	<b>90.8</b>	<b>94.3</b>	<b>65</b>	<b>92.1</b>

observed from Table 6 that the mAP@0.5 of the proposed YOLOv5s-CAM model is 0.6% higher than that of the YOLOv5s model and superior to the other contemporary object detection algorithms in the YOLO series. Even though the precision of YOLOv6 is improved, the rate of false detection is still high. Further the mAP@0.5, mAP@0.5:0.95 and F1-score values of YOLOv5s-CAM are 5.2%, 3.9%, and 4.53% higher than for the YOLOv6s model. On the other hand, the performance of YOLOv7 model is relatively lower, indicating the need for further improvement to achieve more accurate detection of hematoma regions. These results demonstrate the superiority of the YOLOv5s-CAM model among contemporary YOLO models for accurate hematoma detection.

#### D. PERFORMANCE COMPARISON WITH OTHER MODELS USING THE BHX DATASET

To further validate the efficiency of the proposed model, we have compared its performance with other deep learning methods, using the BHX dataset for hematoma detection. The performance analysis of the proposed and the other existing methods are provided in Table 7. It can be seen from Table 7 that the proposed method outperformed the existing models by achieving a higher mAP@0.5 value of 0.943 using the BHX dataset. Further, the comparison results show that the proposed YOLOv5s-CAM model can detect the presence of multiple hematoma subtypes in an individual CT image, with maximum precision and recall values of 0.935 and 0.908. This issue is rarely addressed by previous studies. Further, we have employed ghost

convolution to verify the performance of the YOLOv5s model using ghost convolution for hematoma detection [21]. However, in comparison with the performance of the baseline models as shown in (1) and (2) of Table 5, the mAP@0.5 of the ghost convolution- based YOLOv5s models with SPP and SPPF has dropped down to 0.921 and 0.924 respectively. Moreover, as compared to the YOLOv5x-GCB model [21], the proposed model performs better capture of single or multiple hematoma regions of various subtypes in CT images with an improved mAP@0.5 value of 0.943. In addition, the proposed model was able to perform better discernment of acute and chronic instances of SDH with high precision and recall values as shown in Table 3. Furthermore, the model was able to achieve the highest mAP of 0.943 by splitting the input dataset into the ratio of 60:20:20 for training, validation, and testing. To further validate the performance of the proposed model, the images randomly partitioned in the ratio of 60:20:20 for training, validation, and testing were shuffled to generate three sets of data. Then the mAP achieved by the model using three sets are 0.947, 0.944, and 0.937 respectively and hence the average mAP of the model is  $0.9426 \pm 0.0026$ .

The advantages of the proposed YOLOv5s-CAM model are as follows:

- A deep learning model with cascaded attention modules is developed to detect diverse hematoma lesions of irregular size and shape.
- The model can localize and categorize various hematoma subtypes which include IVH, IPH, EDH, SAH, and SDH respectively.

**TABLE 7. Performance comparison with recent methods.**

Year	Authors	Method	Dataset Used	Classes	Dataset Split	Performance
						P : 0.92
2023	Ertugrul et al. [21]	YOLOv4	BHX dataset with 15,979 images	6	Training and testing - 90:10	R : 0.81 mAP@0.5 : 79.6
						P : 0.92
2023	Kothala et al. [22]	YOLOv5x-GCB	Training and validation: BHX dataset with 21,132 images, Testing: PhysioNet with 2500 images	5	Training and validation- 80:20	R: 0.889 mAP@0.5 : 0.931
						P : 0.77
2023	Ferdi et al. [65]	YOLOv3 + data augmentation	BHX dataset	-	Training and testing - 90:10	R : 0.405 P : 0.949
						P : 0.77
2022	Kothala et al. [66]	DenseNet121	BHX dataset with 26383 CT images	5	Training -20410 Validation - 5973	R : 0.814 mAP@0.5 : 0.931
						P : 0.77
2021	Nemcek et al. [67]	Weakly supervised detection based on multiple instance learning	Training and Validation: RSNA dataset Testing: 254 CT scans from BHX dataset	-	Traning and validation - 97:3	Dice Coefficient : 58.08
						P : 0.77
2021	Karkkainen et al. [68]	Unsupervised Algorithms using mixture model	Training and Validation: RSNA dataset Testing: 198 CT scans from BHX dataset	-		Max Dice Coefficient : 0.814
						P : 0.935
	<b>Proposed Approach</b>	<b>YOLOv5s-CAM</b>	<b>BHX dataset with 15,921 CT images</b>	<b>6</b>	<b>Training, validation, and testing - 60:20:20</b>	<b>R : 0.908</b> <b>mAP@0.5 : 0.943</b>

- The model can discern the acute and chronic cases of SDH, which enables clinicians to perform early interventions and improve patient outcomes.
- The model can localize and categorize multiple hematoma subtypes in a single CT image with the corresponding bounding boxes, class labels, and confidence scores.
- The integration of the stacked window approach with the YOLOv5s-CAM model facilitates better discernment of hematoma regions with improved precision and recall values.

## VI. CONCLUSION

The major objective of the proposed work is to develop a simple, fast, and accurate model to detect single or multiple hematoma regions of irregular size and structure. To balance computational complexity and performance, the standard YOLOv5s model was enhanced by (i) Stacking three grayscale images scaled to different windows in the pre-processing stage to improve the detection of subtle hematoma regions (ii) Integrating cascaded attention module in the neck to strengthen the feature fusion ability and enable better discrimination of ICH subtypes (iii) Replacing SPP with SPPF to improve the receptive

field and multiscale detection. The research study was performed using 15,921 images from the BHX dataset. The experimental results demonstrate that the proposed YOLOv5s-CAM model reached overall precision, recall, F1-score, mAP@0.5 and mAP@0.5:0.95 of 0.935, 0.908, 0.921, 0.943, and 0.65 respectively. Further, the experimental results show that the proposed model improves the recall, mAP@0.5 and mAP@0.5:0.95 by 0.019%, 0.012%, and 0.026% as compared to the YOLOv5x-GCB model.

Limitations are cross-misclassifications between similar-appearing subtypes, and difficulty with accurate detection when hematomas are overlapped by surrounding tissues. The model is implemented using the BHX dataset which consists of a smaller number of positive instances for a few of the hematoma subtypes. Hence, further improvements are required to generalize the performance of the model using other independent datasets. Currently, the model localizes the hematoma regions using the bounding boxes, which are limited to typical rectangular shapes. Hence the model needs further refinement for clinical utilization to estimate hematoma volumes that require precise demarcation of the hematoma regions.

Thus, future research work will focus on enhancing the proposed model by incorporating feature layers in various



locations of the network, optimizing various layers of the YOLOv5s backbone structure, and adopting new feature fusion mechanisms to expand and strengthen the feature information. Further, the model will be trained and tested on other open source datasets to enhance the generalizability and robustness of the model. Moreover, various contour-based models can be applied to generate precise demarcation of the hematoma regions and perform accurate hematoma volume computation.

## ACKNOWLEDGMENT

The authors would like to thank the Manipal Academy of Higher Education, Manipal, for providing the necessary facility to carry out the research. The study was conducted in the Advanced Healthcare Laboratory, Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education.

## REFERENCES

- [1] M. D. Wiles, M. Braganza, H. Edwards, E. Krause, J. Jackson, and F. Tait, "Management of traumatic brain injury in the non-neurosurgical intensive care unit: A narrative review of current evidence," *Anaesthesia*, vol. 78, no. 4, pp. 510–520, Apr. 2023.
- [2] S. L. James, "Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990–2016: A systematic analysis for the global burden of disease study 2016," *Lancet Neurol.*, vol. 18, no. 1, pp. 56–87, 2019.
- [3] C. J. van Asch, M. J. Luitse, G. J. Rinkel, I. van der Tweel, A. Algra, and C. J. Klijn, "Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: A systematic review and meta-analysis," *Lancet Neurol.*, vol. 9, no. 2, pp. 167–176, Feb. 2010.
- [4] L. Gitto and T. E. Richardson, "Anatomopathology and histopathologic changes in chronic subdural hematoma," in *Subdural Hematoma*. Cham, Switzerland: Springer, 2021, pp. 79–93.
- [5] A. Akhaddar, "Review of craniospinal acute, subacute, and chronic subdural hematomas," in *Subdural Hematoma*. Cham, Switzerland: Springer, 2021, pp. 1–24.
- [6] R. Badenes and F. Bilotta, "Neurocritical care for intracranial haemorrhage: A systematic review of recent studies," *Brit. J. Anaesthesia*, vol. 115, pp. ii68–ii74, Dec. 2015.
- [7] V. Vidhya, A. Gudigar, U. Raghavendra, A. Hegde, G. R. Menon, F. Molinari, E. J. Ciaccio, and U. R. Acharya, "Automated detection and screening of traumatic brain injury (TBI) using computed tomography images: A comprehensive review and future perspectives," *Int. J. Environ. Res. Public Health*, vol. 18, no. 12, p. 6499, Jun. 2021.
- [8] A. Shoeibi, M. Khodatars, M. Jafari, N. Ghassemi, P. Moridian, R. Alizadehsani, S. H. Ling, A. Khosravi, H. Alinejad-Rokny, H. K. Lam, M. Fuller-Tyszkiewicz, U. R. Acharya, D. Anderson, Y. Zhang, and J. M. Gorriz, "Diagnosis of brain diseases in fusion of neuroimaging modalities using deep learning: A review," *Inf. Fusion*, vol. 93, pp. 85–117, May 2023.
- [9] S. A. Khan, M. A. Khan, O.-Y. Song, and M. Nazir, "Medical imaging fusion techniques: A survey benchmark analysis, open challenges and recommendations," *J. Med. Imag. Health Informat.*, vol. 10, no. 11, pp. 2523–2531, Nov. 2020.
- [10] V. Vidhya, U. Raghavendra, A. Gudigar, P. Kasula, Y. Chakole, A. Hegde, C. P. Ooi, E. J. Ciaccio, and U. R. Acharya, "Automated intracranial hematoma classification in traumatic brain injury (TBI) patients using meta-heuristic optimization techniques," *Informatics*, vol. 9, no. 1, p. 4, Jan. 2022.
- [11] W. M. Strub, J. L. Leach, T. Tomsick, and A. Vagal, "Overnight preliminary head CT interpretations provided by residents: Locations of misidentified intracranial hemorrhage," *Amer. J. Neuroradiol.*, vol. 28, no. 9, pp. 1679–1682, Oct. 2007.
- [12] N. R. Lal, U. M. Murray, O. P. Eldevik, and J. S. Desmond, "Clinical consequences of misinterpretations of neuroradiologic ct scans by on-call radiology residents," *Amer. J. Neuroradiol.*, vol. 21, no. 1, pp. 124–129, 2000.
- [13] B. Shahangian and H. Pourghassem, "Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 217–232, 2016.
- [14] I. Kumar, C. Bhatt, and K. U. Singh, "Entropy based automatic unsupervised brain intracranial hemorrhage segmentation using CT images," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 6, pp. 2589–2600, Jun. 2022.
- [15] D. M. Alawad, A. Mishra, and M. T. Hoque, "AIBH: Accurate identification of brain hemorrhage using genetic algorithm based feature selection and stacking," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 2, pp. 56–77, Apr. 2020.
- [16] P. Inkeaw, S. Angkurawaranon, P. Khumrin, N. Inmutto, P. Traisathit, J. Chaijaruanich, C. Angkurawaranon, and I. Chitapanarux, "Automatic hemorrhage segmentation on head CT scan for traumatic brain injury using 3D deep learning model," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105530.
- [17] A. Phaphuangwittayakul, Y. Guo, F. Ying, A. Y. Dawod, S. Angkurawaranon, and C. Angkurawaranon, "An optimal deep learning framework for multi-type hemorrhagic lesions detection and quantification in head CT images for traumatic brain injury," *Int. J. Speech Technol.*, vol. 52, no. 7, pp. 7320–7338, May 2022.
- [18] M. Ganeshkumar, V. Ravi, V. Sowmya, E. Gopalakrishnan, K. Soman, and C. Chakraborty, "Identification of intracranial haemorrhage (ICH) using ResNet with data augmentation using CycleGAN and ICH segmentation using SegAN," *Multimedia Tools Appl.*, vol. 81, no. 25, pp. 36257–36273, Oct. 2022.
- [19] Y. Ma, F. Ren, W. Li, N. Yu, D. Zhang, Y. Li, and M. Ke, "IHA-Net: An automatic segmentation framework for computer-tomography of tiny intracerebral hemorrhage based on improved attention U-Net," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104320.
- [20] Ö. F. Ertuğrul and M. F. Akil, "Detecting hemorrhage types and bounding box of hemorrhage by deep learning," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103085.
- [21] L. P. Kothala, P. Jonnala, and S. R. Guntur, "Localization of mixed intracranial hemorrhages by using a ghost convolution-based YOLO network," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104378.
- [22] A. Gudigar, U. Raghavendra, A. Hegde, M. Kalyani, E. J. Ciaccio, and U. Rajendra Acharya, "Brain pathology identification using computer aided diagnostic tool: A systematic review," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105205.
- [23] Z. Ji, Y. Wu, X. Zeng, Y. An, L. Zhao, Z. Wang, and I. Ganchev, "Lung nodule detection in medical images based on improved YOLOv5s," *IEEE Access*, vol. 11, pp. 76371–76387, 2023.
- [24] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [25] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1137–1149.
- [27] M. E. Salman, G. Çakirsoy Çakar, J. Azimjonov, M. Kösem, and I. H. Cedimoğlu, "Automated prostate cancer grading and diagnosis system using deep learning-based YOLO object detection algorithm," *Expert Syst. Appl.*, vol. 201, Sep. 2022, Art. no. 117148.
- [28] N. Lv, J. Xiao, and Y. Qiao, "Object detection algorithm for surface defects based on a novel YOLOv3 model," *Processes*, vol. 10, no. 4, p. 701, Apr. 2022.
- [29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [30] *Ultralytics, YOLOv5*. Accessed: Jan. 16, 2023. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [31] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, and L. Wang, "Image thresholding improves 3-Dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans," *Sensors*, vol. 19, no. 9, p. 2167, May 2019.
- [32] W. Kuo, C. Häne, P. Mukherjee, J. Malik, and E. L. Yuh, "Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 45, pp. 22737–22745, Nov. 2019.



- [33] J. Cho, K.-S. Park, M. Karki, E. Lee, S. Ko, J. K. Kim, D. Lee, J. Choe, J. Son, M. Kim, S. Lee, J. Lee, C. Yoon, and S. Park, "Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models," *J. Digit. Imag.*, vol. 32, no. 3, pp. 450–461, Jun. 2019.
- [34] T. Shi, W. Zhu, and Y. Su, "Improved light-weight target detection method based on YOLOv5," *IEEE Access*, vol. 11, pp. 38604–38613, 2023.
- [35] A. Arshaghi, M. Ashourin, and L. Ghabeli, "Detection and classification of potato diseases potato using a new convolution neural network architecture," *Traitement du Signal*, vol. 38, no. 6, pp. 1783–1791, Dec. 2021.
- [36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [37] N. Zaghari, M. Fathy, S. M. Jameii, and M. Shahverdy, "The improvement in obstacle detection in autonomous vehicles using YOLO non-maximum suppression fuzzy algorithm," *J. Supercomput.*, vol. 77, no. 11, pp. 13421–13446, Nov. 2021.
- [38] H. Zhang, M. Tian, G. Shao, J. Cheng, and J. Liu, "Target detection of forward-looking sonar image based on improved YOLOv5," *IEEE Access*, vol. 10, pp. 18023–18034, 2022.
- [39] M. Karki, J. Cho, E. Lee, M.-H. Hahm, S.-Y. Yoon, M. Kim, J.-Y. Ahn, J. Son, S.-H. Park, K.-H. Kim, and S. Park, "CT window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings," *Artif. Intell. Med.*, vol. 106, Jun. 2020, Art. no. 101850.
- [40] V. Vidhya, A. Gudigar, R. U. S. Basak, S. Mallappa, A. Hegde, and G. Menon, "An efficient detection of intracranial hematoma using window-based stacking and YOLOv5 framework," in *Proc. 2nd Int. Conf. Advancements Electr., Electron., Commun., Comput. Autom. (ICAECA)*, Jun. 2023, pp. 1–6.
- [41] T.-H.-Y. Le, A.-C. Phan, H.-P. Cao, and T.-C. Phan, "Automatic identification of intracranial hemorrhage on CT/MRI image using meta-architectures improved from region-based CNN," in *Optimization of Complex Systems: Theory, Models, Algorithms and Applications*. Cham, Switzerland: Springer, 2020, pp. 740–750.
- [42] Z. Xue, S. Antani, L. R. Long, D. Demner-Fushman, and G. R. Thoma, "Window classification of brain CT images in biomedical articles," in *Proc. AMIA Annu. Symp.*, 2012, p. 1023.
- [43] X. Lu and W. Song, "Improved YOLOv5s model for vehicle detection and recognition," in *Proc. Int. Conf. Intell. Comput.* Cham, Switzerland: Springer, 2022, pp. 423–434.
- [44] G. Brauwers and F. Frasincar, "A general survey on attention mechanisms in deep learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 4, pp. 3279–3298, Apr. 2023.
- [45] G. Liu, Y. Hu, Z. Chen, J. Guo, and P. Ni, "Lightweight object detection algorithm for robots with improved YOLOv5," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106217.
- [46] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
- [47] L. Zhu, X. Geng, Z. Li, and C. Liu, "Improving YOLOv5 with attention mechanism for detecting boulders from planetary images," *Remote Sens.*, vol. 13, no. 18, p. 3776, Sep. 2021.
- [48] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 3–19.
- [49] P. Li, T. Han, Y. Ren, P. Xu, and H. Yu, "Improved YOLOv4-tiny based on attention mechanism for skin detection," *PeerJ Comput. Sci.*, vol. 9, Mar. 2023, Art. no. e1288.
- [50] J. Wang, J. Yu, and Z. He, "DECA: A novel multi-scale efficient channel attention module for object detection in real-life fire images," *Int. J. Speech Technol.*, vol. 52, no. 2, pp. 1362–1375, Jan. 2022.
- [51] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Netw.*, vol. 107, pp. 3–11, Nov. 2018.
- [52] E. P. Reis, F. Nascimento, M. Aranha, F. M. Seicol, B. Machado, M. Felix, A. Stein, and E. Amaro, "Brain hemorrhage extended (bhx): Bounding box extrapolation from thick to thin slice ct images," *PhysioNet*, vol. 101, no. 23, p. e215, 2020.
- [53] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Development and validation of deep learning algorithms for detection of critical findings in head CT scans," 2018, *arXiv:1803.05854*.
- [54] U. Igbaseimokumo, *Brain CT Scans in Clinical Practice*. Cham, Switzerland: Springer, 2009.
- [55] B. Lee and A. Newberg, "Neuroimaging in traumatic brain imaging," *NeuroRX*, vol. 2, no. 2, pp. 372–383, Apr. 2005.
- [56] J. Muschelli, E. M. Sweeney, N. L. Ullman, P. Vespa, D. F. Hanley, and C. M. Crainiceanu, "PitcHPERFeCT: Primary intracranial hemorrhage probability estimation using random forests on CT," *NeuroImage, Clin.*, vol. 14, pp. 379–390, Jan. 2017.
- [57] S. Saponara and A. Elhanashi, "Impact of image resizing on deep learning detectors for training time and model performance," in *Proc. Int. Conf. Appl. Electron. Pervading Ind., Environ. Soc.* Cham, Switzerland: Springer, 2021, pp. 10–17.
- [58] O. Rukundo, "Effects of image size on deep learning," *Electronics*, vol. 12, no. 4, p. 985, Feb. 2023.
- [59] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi, and A. K. Bhoi, "Statistical analysis of design aspects of various YOLO-based deep learning models for object detection," *Int. J. Comput. Intell. Syst.*, vol. 16, no. 1, p. 126, Aug. 2023.
- [60] I. S. Isa, M. S. A. Rosli, U. K. Yusof, M. I. F. Maruzuki, and S. N. Sulaiman, "Optimizing the hyperparameter tuning of YOLOv5 for underwater detection," *IEEE Access*, vol. 10, pp. 52818–52831, 2022.
- [61] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jul. 2020, pp. 237–242.
- [62] Z. Wang, Y. Wu, L. Yang, A. Thirunavukarasu, C. Evison, and Y. Zhao, "Fast personal protective equipment detection for real construction sites using deep learning approaches," *Sensors*, vol. 21, no. 10, p. 3478, May 2021.
- [63] T. Xu, X. Gao, Y. Yang, L. Xu, J. Xu, and Y. Wang, "Construction of a semantic segmentation network for the overhead catenary system point cloud based on multi-scale feature fusion," *Remote Sens.*, vol. 14, no. 12, p. 2768, Jun. 2022.
- [64] J. Verma and J. Verma, "One-way ANOVA: Comparing means of more than two samples," in *Data Analysis in Management With SPSS Software*. India: Springer, 2013, pp. 221–254.
- [65] A. Ferdi, S. Benierbah, and Y. Ferdi, "YOLOv3-based intracranial hemorrhage localization from CT images," in *Proc. 13th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, Mar. 2023, pp. 1–6.
- [66] L. P. Kothala and S. R. Guntur, "Multi-class classification of intracranial hemorrhages in a 3-Channel CT image by using a transfer learning based DenseNet121 model," in *Proc. Int. Conf. Smart Gener. Comput., Commun. Netw. (SMART GENCON)*, Dec. 2022, pp. 1–5.
- [67] J. Nemcek, T. Vicar, and R. Jakubicek, "Weakly supervised deep learning-based intracranial hemorrhage localization," 2021, *arXiv:2105.00781*.
- [68] K. Kärkkäinen, S. Fazeli, and M. Sarrafzadeh, "Unsupervised acute intracranial hemorrhage segmentation with mixture models," in *Proc. IEEE 9th Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2021, pp. 120–129.



**V. VIDHYA** received the M.E. degree. She is currently pursuing the Ph.D. degree with the Manipal Academy of Higher Education, India. She is also a Faculty Member of the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal, India. Her research interests include image processing and medical image analysis.



**U. RAGHAVENDRA** received the Ph.D. degree from the Manipal Academy of Higher Education, India. He is currently a Faculty Member of the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal, India. He has published several papers in refereed international SCI-IF journals and international conference proceedings. He has a patent to his credit and received an invention award from intellectual ventures in USA, for his innovations, in 2014. His research interests include 3D computer vision, image processing, and medical image analysis. For more information visit the link (<https://scholar.google.co.in/citations?user=3nzcDREAAA&hl=en>).



**AJAY HEGDE** is currently a Consultant in neurosurgeon with Manipal Hospitals, Bengaluru. He has completed his training in neurosurgery from the Kasturba Medical College, Manipal, and pursued further training in general and functional neurosurgery with the Institute of Neurological Sciences, Glasgow, U.K. With an experience of over 2500 neurosurgical procedures, he is trained to handle both routine and complex procedures, with a strong focus on patient safety and outcome. He has undertaken several projects in digital health, database structuring, and application of artificial intelligence in healthcare. He has coauthored several works on AI-based classification of images.



**ANJAN GUDIGAR** received the Ph.D. degree from the Manipal Academy of Higher Education, India. He is currently a Faculty Member of the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal, India. He has published several research papers in international conferences and journals. His research interests include image processing, medical image analysis, and pattern recognition. For more information visit the link (<http://scholar.google.co.in/citations?user=qoe6EvsAAA&hl=en>).



**GIRISH R. MENON** is currently a dedicated Neuro Surgeon with more than 23 years of experience in medical and hospital administration. He is also the Head of the Department of Neurosurgery, Kasturba Medical College, Manipal. He has driven the implementation of novel AI-based algorithms in the treatment of intracerebral hemorrhage in clinical practice. His research interests include cerebrovascular and neuroendoscopic surgeries.



**SUDIPTA BASAK** is currently pursuing the B.Tech. degree with the Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. His research interests include medical image analysis using machine learning and deep learning.



**PRABAL DATTA BARUA** received the Ph.D. degree in information systems from the University of Southern Queensland (UniSQ), Australia. Currently, he is the Academic Dean of the Australian International Institute of Higher Education and an Adjunct Professor with UniSQ, where he is also a member of the ICT Industry Advisory Board, the ICT Programs Leadership & Governance Committee, and the Business in AI Steering Committee. He is also an accredited Higher Degree by Research (H.D.R.) Supervisor with UniSQ. He is also an Adjunct Professor with Taylor's University, Malaysia, and an Adjunct Professor with the University of New England (UNE), where he is also an accredited H.D.R. Supervisor. He is also a Visiting Professor with the SRM University of Science and Technology, India. He is also the CEO of Cogninet Australia, which conducts research in medical and educational informatics using artificial intelligence (AI). He is also a Honorary Professional/Industry Fellow with the University of Technology Sydney (UTS). He is the prolific author with over 100 publications in top-tier academic journals and has contributed as a guest editor of many international journals. He has secured over A\$3M in government and industry research grants for conducting cutting-edge research in applying artificial intelligence in health informatics, education analytics, and ICT for business transformation.



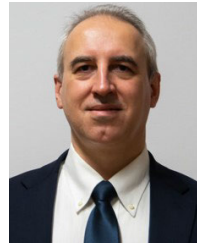
**SANKALP MALLAPPA** is currently pursuing the B.Tech. degree with the Department of Electronics and Instrumentation Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal. His research interests include medical image analysis using machine learning and deep learning.



**MASSIMO SALVI** received the master's degree in biomedical engineering from Politecnico di Torino, Turin, Italy, in 2014, and the Ph.D. degree, in 2018, with a thesis titled "Modeling and Interpretation of Physiopathological Processes Using Multimodal and Multiscale Medical Imaging." In addition to his work in biomedical signal and image processing, he is also actively engaged in the field of artificial intelligence and deep learning. He has experience in the development of hybrid techniques that combine AI-based methods and mathematical-statistical techniques to address complex problems in biomedical research. His research interests include the development of automated solutions for fluorescence and optical microscopy, with a particular emphasis on the analysis and interpretation of cellular and tissue images.



**EDWARD J. CIACCIO** received the M.S. and Ph.D. degrees from the Department of Biomedical Engineering, Rutgers University. He is currently a Senior Research Scientist with the Division of Cardiology, Department of Medicine, and the Celiac Disease Center, Columbia University, and also a Honorary Principal Research Fellow with the Department of Medicine, Division of Cardiology, Imperial College London. He has published over 200 peer-reviewed articles on such topics as biomedical signal processing of heart electrograms and image processing of villous atrophy in celiac disease patients. In computational biology, he has developed biophysical models of activation wavefront propagation for ventricular tachycardia and for atrial fibrillation. His work received a Paper of the Year Award from Heart Rhythm journal, one of the leading publications in cardiac electrophysiology, an Established Investigator Award from the American Heart Association, the Whitaker Foundation Research Award, and four U.S. patents. He is the former Editor-in-Chief of *Computers in Biology and Medicine* (2013–2021) and the founding Editor-in-Chief of *Informatics in Medicine Unlocked* journal.



**FILIPPO MOLINARI** (Senior Member, IEEE) is currently a Full Professor with Politecnico di Torino, where he leads research in several areas of biomedical engineering. His primary research interests include biomedical signal processing, image processing, and ultrasound technology, with a focus on developing advanced diagnostic tools for a range of medical applications. In addition to his work in these areas, his research interests include the non-invasive characterization of tumor vascularization, neuroimaging for the advanced assessment of neurodegenerative disorders, and neurovascular and metabolic assessment of cerebral autoregulation.



**U. RAJENDRA ACHARYA** (Senior Member, IEEE) received the Ph.D., D.Eng., and D.Sc. degrees. He is currently a Professor in artificial intelligence in healthcare with the University of Southern Queensland, Australia; a Distinguished Professor with the International Research Organization for Advanced Science and Technology, Kumamoto University, Japan; and an Adjunct Professor with the University of Malaya, Malaysia, and Asia University, Taiwan. His research interests include biomedical imaging and signal processing, data mining, visualization, and the applications of biophysics for better healthcare design and delivery. His funded research has accrued cumulative grants exceeding six million Singapore dollars. He has authored over 600 publications, including 550 in refereed international journals, 42 in international conference proceedings, and 17 books. He has received more than 70,000 citations on Google Scholar (with an H-index of 132). He has been ranked in the top 1% of the highly cited researchers for the last seven consecutive years (2016–2022) in computer science, according to the Essential Science Indicators of Thomson. He is on the editorial boards of many journals and has served as a Guest Editor for several AI-related issues.

...