Biomedical Signal Processing and Control 80 (2023) 104378

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Localization of mixed intracranial hemorrhages by using a ghost convolution-based YOLO network

Lakshmi Prasanna Kothala [a], Prathiba Jonnala [a], Sitaramanjaneya Reddy Guntur [b,*]

[a] Department of Electronics and Communication Engineering, Vignan's Foundation for Science, Technology, and Research, Guntur, Andhra Pradesh 522213, India
[b] Department of Biomedical Engineering, Vignan's Foundation for Science, Technology, and Research, Guntur, Andhra Pradesh 522213, India

ARTICLE INFO

Keywords:
Computer Tomography
Deep Learning
Ghost Convolution
Mixed Intracranial Hemorrhage
YOLO Models
Ghost convolution-based YOLO (YOLOv5x-GCB)

ABSTRACT

Intracranial hemorrhage (ICH) is a serious medical condition that must be diagnosed in a stipulated time through computed tomography (CT) imaging modality. However, the neurologist must initially confirm the specific type of hemorrhage to prescribe an effective treatment. Although conventional image processing and convolution-based deep learning models can effectively perform multiclass classification tasks, they fail to classify if a CT input image contains multiple hemorrhages in a single slice and takes a lot of time to make the final predictions. To overcome these two difficulties, we proposed a novel YOLOv5x-GCB model that can be able to detect multiple hemorrhages with limited resources by employing a ghost convolution process. The advantage of ghost convolution is that it produces the same number of feature maps as vanilla convolution while using less expensive linear operations. Another feature of the proposed model is that it uses the mosaic augmentation technique throughout the training to improve the accuracy of mixed hemorrhage detection. A brain hemorrhage extended dataset containing 21,132 slices from 205 positive patients was used in training and validating the proposed model. To test the robustness of the proposed model, we created a separate dataset with the existing segmentation data, which are available in PhysioNet. As a result, the proposed model achieved an overall precision, recall, F1- score, and mean average precision of 92.1%, 88.9%, 90%, and 93.1%, respectively. In addition to these metrics, other parameters were used in evaluating the proposed model and checking its lightweight capability in terms of memory size and computational time. Results showed that our proposed model can be used in real-time clinical diagnosis by using either embedded devices or cloud services.

1. Introduction

Intracranial hemorrhage (ICH) is a dangerous condition with a 40 % mortality rate within one month. It refers to bleeding, which occurs inside the skull for one of two reasons: it can occur spontaneously or can be caused by trauma. Sudden ICH can be caused by a variety of diseases, whereas traumatic ICH can only occur in the context of trauma [1–3]. Patients who take anticoagulants are at a high risk of developing ICH. According to the position of bleeding, ICH is classified into five different subtypes: cerebral parenchyma hemorrhage (CPH) or intraparenchymal hemorrhage (IPH), epidural hemorrhage (EDH), intraventricular hemorrhage (IVH), subarachnoid hemorrhage, and subdural hemorrhage (SDH) [4,5]. Currently, the clinical technique for diagnosing ICH is analyzing computed tomography (CT) images by radiologists. However, the availability of skilled radiologists and their presence in rural areas is a difficult task because of the lack of resources in these locations [6,7]. Sometimes there is a possibility of mixed hemorrhages in a single patient itself due to multiple fractures in the brain during injury as shown in Fig. 1. As a result, there is a need to automate the ICH detection process.

Recently, deep learning (DL) techniques have been applied to medical images as they produce tremendous results in the segmentation and classification of ICH. DL is a subdivision of artificial intelligence and uses the concept of the human brain's decision-making process to build algorithms. DL algorithms have attracted considerable interest as they use the higher dimensionality of training data and have marked higher detection accuracy than traditional methods. One additional advantage of DL models is that they extract features automatically from raw input data, and the main idea behind the working of DL is a gradient concept. Several DL models have been developed for the treatment of ICH in its early stage [8–9]. However, the main drawback of these methods is that they can classify different types of ICH only if a patient has only one type of ICH, that is, they cannot detect mixed hemorrhage cases [10,11].

* Corresponding author at: Department of Biomedical Engineering, Vignan's Foundation for Science, Technology and Research, Vadlamudi, Guntur-522213, India.
E-mail address: gunturu.ram@hotmail.com (S.R. Guntur).

https://doi.org/10.1016/j.bspc.2022.104378
Received 15 May 2022; Received in revised form 19 September 2022; Accepted 30 October 2022

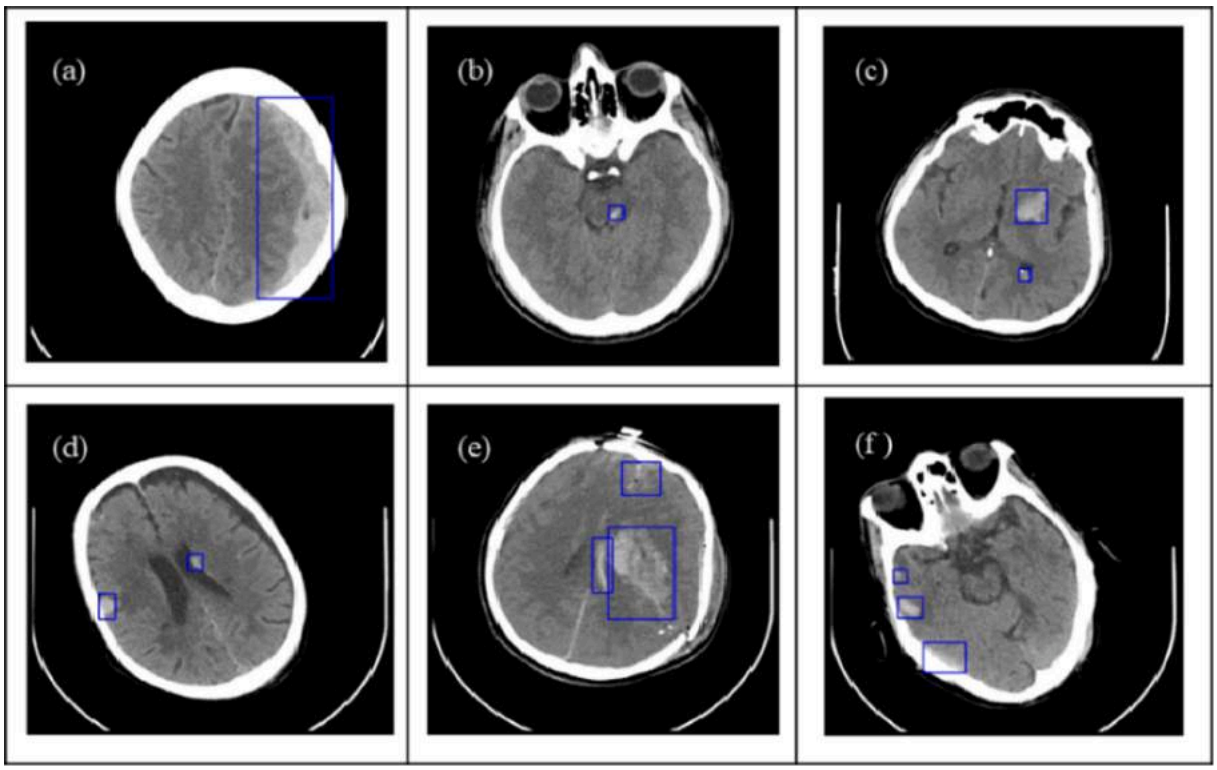

Fig. 1. Illustration of CT axial slice scans demonstrating single and multiple ICH cases in a given single slice: (a) and (b) only one hemorrhage, (c) and (d) two types of hemorrhages, (e) and (f) three types of hemorrhages.

To overcome this limitation, object detection algorithms are used in classifying mixed ICH in a CT image. The object detection algorithm has a two-step process: classifying each object in a given image along with its localization. Object detection algorithms are divided into two categories, namely, classification, and regression-based algorithms [12]. The first category divides an input image into several regions known as region proposals. A convolution process is then applied to each region individually for the acquisition of feature maps from each region. Some popular algorithms in this category include region-based convolutional neural networks (RCNNs), fast RCNN, faster RCNN, and mask RCNN. However, they take a long time to converge because they must consider predictions from each region to obtain the final predictions. In the second category, convolution is applied over the entire image at a time to obtain the final prediction. The best examples in this category are YOLO and single-shot detector (SSD) [13–16]. Owing to technological advancements, various version of YOLO has been released, such as YOLOv2, YOLOv3, YOLOv4, and YOLOv5, to improve detection accuracy [17–20].

The main motivation behind this research is to create a model that can localize mixed hemorrhages with limited resources. To achieve this, we proposed a novel architecture named YOLOv5x-GCB, which was developed on the basis of a standard YOLOv5x and ghost convolution (GC) module. Here, YOLOv5x was used in the localization of mixed ICH, while the GC module was used to reduce the complexity of the proposed model. The main contributions of the proposed work are summarized as follows:

(i) The proposed YOLOv5x-GCB model is trained using an advanced data augmentation technique called Mosaic to improve detection accuracy while localizing the mixed ICH in a given CT image.

(ii) The proposed model is lightweight (requires less memory) and fast as it has a high frame per second (FPS), which was achieved by employing a GC mechanism rather than a traditional vanilla convolution method.

(iii) The performance of the proposed model was compared with other state-of-the-art models to determine its robustness. To make this comparison, we considered four metrics: recall, precision, F1 score, and mean average precision (mAP) at different thresholds. Apart from the above-mentioned metrics, we considered other parameters such as FPS, iterations per second (IPS), total execution time, and the number of trainable parameters required while executing the proposed model to check the speed and memory required to deploy the proposed model in the cloud.

The different sections of this research paper are organized as follows: Section II presents a literature review of traditional methods and current DL models for detecting ICH. Section III describes the benchmark methodology used in detecting mixed ICH. Section IV presents the dataset description, training process, and experimental results. Section V provides a related discussion and a comparison to determine the best model. Finally, the limitations and future scope of the proposed model are discussed.

2. Related works

Currently, research on ICH detection, segmentation, and classification using CT scans is divided into two categories related to traditional algorithms and DL models.

2.1. Traditional ICH detection

In traditional methods, that is, prior to the existence of DL algorithms, Chan et al. (2007) used image thresholding and morphological operations to remove the skull area and used a median filter a preprocessing technique [5]. A hat transformation technique was then used in extracting hemorrhage candidates, and a knowledge-based classifier was used to distinguish the true and false hemorrhage pixels. A validation sensitivity and specificity of 100 % and 84 % were achieved, respectively. However, owing to the overlapping region of the hemorrhage with gray matter, the above process may not always produce

positive results. To solve this problem, Bhadauria et al. (2014) proposed a method called a fuzzy-based level set to segment a hemorrhage area from a normal one [21]. They initially used an unsupervised clustering algorithm, which divided input images into groups according to similarity, followed by a region-based active contour to segment the hemorrhages. Their algorithm achieved sensitivity, specificity, Jaccard index (JI), and dice coefficient (DC) of 79.4 %, 99.4 %, 78.2 %, and 87.4 %, respectively. Alawad et al. (2020) used Otsu's thresholding method as a preprocessing technique, and the features were extracted manually from a region of interest [22]. A genetic algorithm was then used to select the best features, which were fed into three individual classifiers to determine the best model. Accuracy, precision, recall, and F1-score of 99.5 %, 99 %, 98.9 %, and 98.9 % were achieved, respectively.

2.2. Deep Learning-based ICH detection

Most researchers have used DL algorithms effectively on brain images to deal with problems, such as hemorrhages and tumors. With regard to hemorrhage, they primarily focus on segmentation, binary classification, and multiclass classification. In general, the size of a hemorrhage determines the severity of a patient. Thus, segmentation tasks play a key role in ICH diagnosis. Hssayeni et al. (2020) proposed a patch-based U-Net model and achieved a recall, specificity, and JI of 97.2 %, 50.4 %, and 0.21, respectively [23]. The obtained recall was higher than the proposed model because they considered the JI value as 0.21, which is similar to the IOU threshold in the proposed model with the consideration of 0.5. Ali Arab et al. (2020) utilized the deep supervision concept in an existing CNN and obtained precision, recall, and F1-score of 85 %, 83 %, and 84 %, respectively. These values were lower than those of the proposed model [24].

Ground truth annotations are required to segment the exact area of hemorrhage, which are provided by radiologists, and obtaining them is a time-consuming process and expensive. Thus, Wang et al. (2020) proposed a new technique called semi-supervised learning, which uses labeled and unlabeled data to train the U-Net model [25]. Here, the author used an extremely limited number of labeled data and huge unlabeled data. Compared with the results of supervised learning alone, semi-supervised learning had higher DC (67 %) and JI (50 %). However, the availability of medical images in the public domain is limited. Wu et al. (2020) proposed a new technique called generative adversarial networks (GANs) to generate synthetic data from original data [26]. GANs use two modules named generator and discriminator are used to accomplish the task and with this model, they achieved precision and recall of 72.1 % and 52.3 %, respectively.

In the next work, Ganeshkumar et al. (2022) proposed a ResNet model for binary classification, and then to segment the exact location, a novel model called SegAN is used with CycleGAN data augmentation as a preprocessing technique [27]. The precision, recall, and F1 score were 94 %, 89 %, and 91 %, respectively. However, the subtypes of ICH were not considered. Qiu et al. (2019) proposed a U-Net model with transfer learning, where an encoder was pretrained with the ResNet model [28]. Precision and recall of 93.5 % and 95 % were achieved, respectively. However, mixed hemorrhage cases were not considered. Cho et al. (2019) proposed a transfer learning model constructed with a pair of CNN and FCN [29]. VGG16 was used in detecting the presence of ICH, whereas FCN was used to segment the delineation of ICH. Sensitivity and specificity of 97.9 %, and 98.7 %, respectively were achieved during detection, and precision and recall of 80.1 %, and 82.1 %, respectively, were achieved during segmentation. Mixed hemorrhages were segmented, but the results were lower than those of the proposed model. Finally, to calculate the volumes of ischemic and hemorrhagic stroke, Kuang et al. (2019) proposed a modified U-Net architecture that uses convex optimization [30]. The main advantage of our proposed model over segmentation models is we achieved the highest performance metrics by considering the IOU threshold as 0.5. Moreover, we detected mixed hemorrhages while most segmentation works concentrated on a single hemorrhage in a given CT.

In the classification models, Thay et al. (2018) proposed a random forest classifier and achieved precision and recall of 99 % and 98.8 %, respectively [31]. Zhou et al. (2022) proposed two models based on transfer learning and achieved recall values of 87.4 for the Resnet model and 80.2 for the DenseNet-121 model [32]. Vrbancic et al. (2019) used the gray wolf optimization technique to tune the hyperparameters before transfer learning and achieved precision, recall, and F1 score of 90 %. 93 %, and 91 %, respectively [33]. Chen et al. (2022) proposed three models based on the concept of transfer learning and achieved high accuracy in predicting whether a patient is having an ICH stroke, infarct, or other [34]. Despite their high detection accuracy, these methods are limited to binary classification. To achieve the multiclass classification, Chilamkurthy et al. (2018) proposed a ResNet18 model with five fully connected layers at the output layer [35]. Furthermore, each slice was labeled using a natural language processing algorithm, and the predictions had sensitivity and specificity of 92 % and 70 %, respectively. Sage et al. (2020) used two identical ResNet50 models with different preprocessing techniques [36]. Then, the feature maps from the two modules were concatenated to produce a large number of features. Finally, the classification results were compared with two different classifiers and the F1 score was 88.2 %. To obtain slice-wise information, Ye et al (2019) used a recurrent neural network (RNN) in conjunction with CNN [37]. A 3D CT was used in determining whether a patient had a hemorrhage or not by using CNN, and its subtype was classified with an RNN. The sensitivity and F1 scores were 83.1 % and 85.5 %, respectively.

Recently, a few papers on object detection algorithms have been published. Ferlin et al (2021) proposed a 2D faster R-CNN network and obtained precision, recall, and F1 score of 89.7 %, 92.6 %, and 90.8 %, respectively [38]. Le THY et al. (2019) proposed two architectures and obtained higher precision (90.5 %), recall (82.6 %), and F1 score (86.45) for the R-FCN model [39]. Li et al (2021) proposed an SSD architecture with a pretrained VGG and achieved precision and F1 scores of 79.7 and 84.5 %, respectively [40]. Al-masni et al (2020) proposed a combination of YOLO and 3D CNN and achieve precision and F1 scores of 67.2 and 77.6 %, respectively [41]. Myung et al. (2021) proposed a YOLOv2 network based on a pre-trained ResNet-50 backbone, and the values improved to 79.7 % and 84.5 % [42].

Finally, Ertugrul et al. (2022) used the YOLOv4 architecture to draw a bounding box around a hemorrhage region [43]. By training all of the different types of hemorrhages individually, they achieved an overall average percentage of 92.8 %, 93.8 %, 91.8 %, and 90.6 % for F1 score, precision, recall, and mAP, respectively. However, these values were reduced to 86 %, 92 %, 81 %, and 79.6 % when all types were considered simultaneously. These results suggest that model performance can be improved in the presence of mixed hemorrhages. To overcome this limitation a novel YOLOv5x–GCB model was proposed, which will improve the accuracy to great extent even with limited resources.

3. Proposed methodology

This research aims to develop a model that can detect mixed hemorrhages with high accuracy and limited resources using the bounding box concept. YOLO has become a subject of interest in the computer vision field because of its fast-computing capability [17]. YOLO creates a 19×19 grid on an input image and calculates an output value for each grid with the parameters [*Pc*, Bx, By, Bh, Bw, CS1, CS2, CS3, CS4, and CS5], where *Pc* denotes the probability of a class value in a specific grid; Bx, By, Bh, and Bw denote the dimensions of the bounding box; and CS1, CS2, CS3, CS4, and CS5 denote the confidence scores of the five different class labels. Sometimes, more than one box is predicted around a single hemorrhage. In such a case, an algorithm known as non-max suppression is used to delete all duplicate boxes with a *Pc* less than 0.6. By contrast, if a grid itself contains multiple hemorrhages, a fixed number of anchor boxes with constant dimensions are tiled over each grid, and

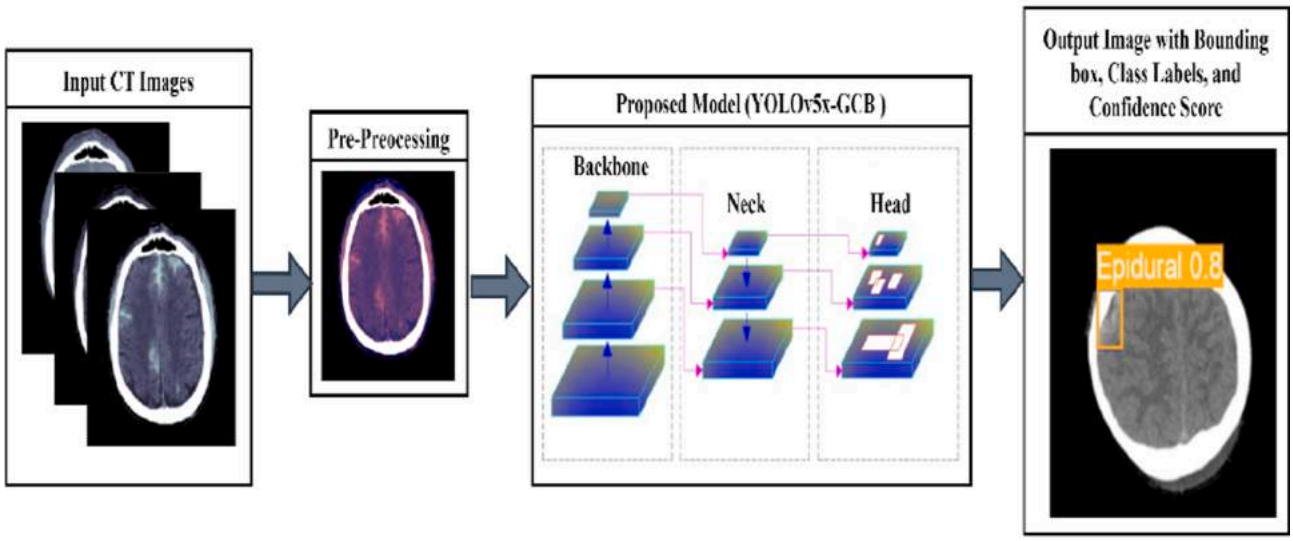

Fig. 2. Block diagram representing the total methodology for detecting mixed ICH in a given input CT image by drawing a rectangular bounding box around the hemorrhage along with its label name and a confidence score value.

Table 1
Comparison of YOLOv5 models: YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and proposed (YOLOv5x-GCB) model in terms of model depth and width and the count of C3 and C3Ghost module with respect to each model.

Parameter	YOLO v5s	YOLO v5s-GCB	YOLOv5m	YOLO v5m-GCB	YOLOv5l	YOLO v5l-GCB	YOLOv5x	Proposed Model
Depth Multiple	0.33	0.33	0.67	0.67	1.0	1.0	1.33	1.33
Width Multiple	0.50	0.50	0.75	0.75	1.0	1.0	1.25	1.25
Count of C3 and C3Ghost in YOLOv5 backbone	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4
	C3-2	C3Ghost-2	C3-4	C3Ghost-4	C3-6	C3Ghost-6	C3-8	C3Ghost-8
	C3-3	C3Ghost-3	C3-6	C3Ghost-6	C3-9	C3Ghost-9	C3-12	C3Ghost-12
	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4
Count of C3 and C3Ghost in YOLOv5 neck	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4
	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4
	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4
	C3-1	C3Ghost-1	C3-2	C3Ghost-2	C3-3	C3Ghost-3	C3-4	C3Ghost-4

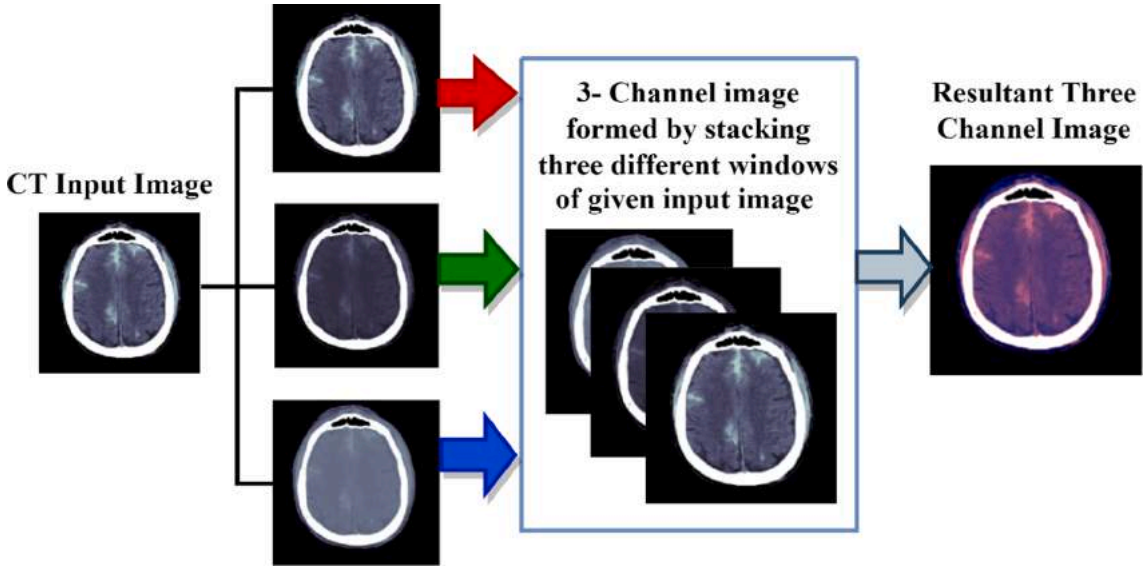

Fig. 3. Formation of a three-channel RGB image from a single CT slice by exploiting three different window settings: brain, subdural, and soft tissue. The resultant three-channel images will be fed as inputs to the proposed model.

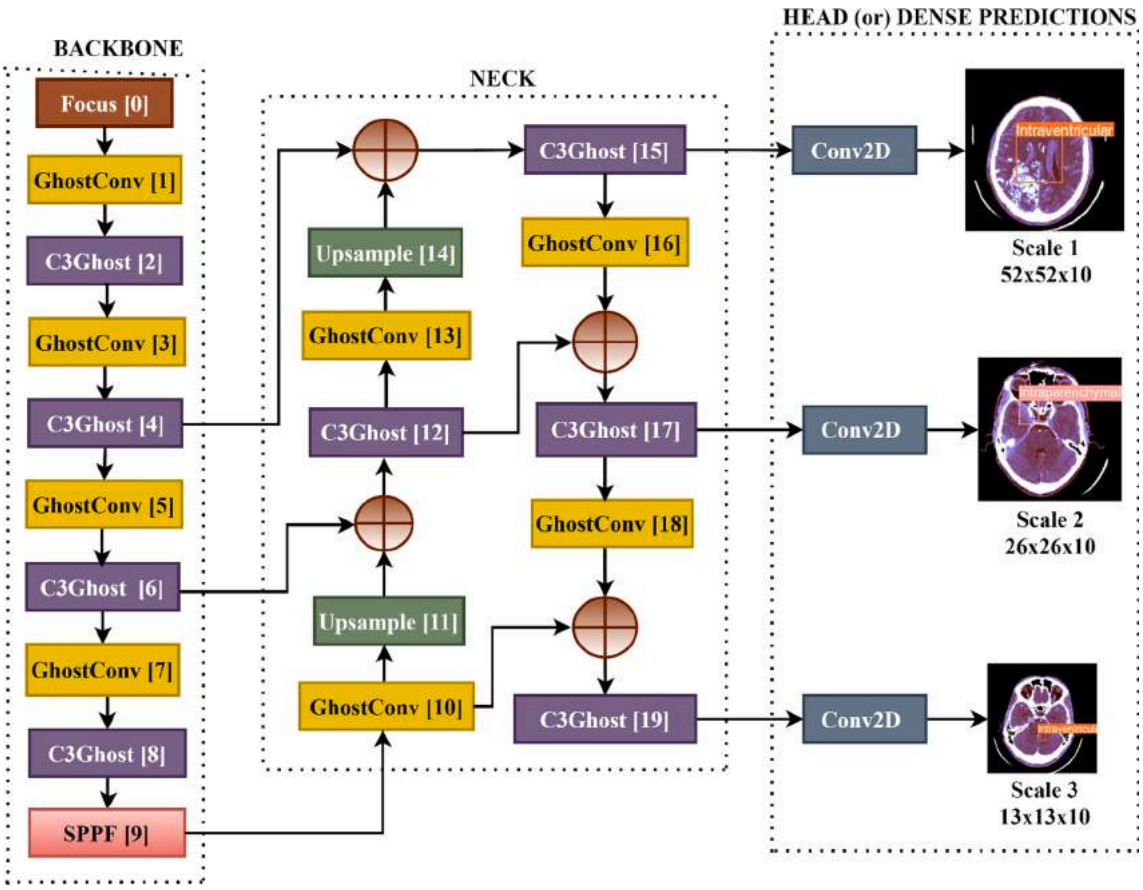

Fig. 4. Proposed YOLOv5x-GCB architecture and its all-internal modules.

the rest of the process is similar. Finally, the performance of YOLO is measured by assigning a threshold to the IOU value. The entire flow of the proposed methodology is shown in Fig. 2.

3.1. Proposed model (YOLOv5x-GCB)

YOLO is one of the popular single-stage models used for object detection. Conventional YOLOv5 can be implemented in four ways: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These models are designed by varying the number of C3 (bottleneck CSP with three convolutions) modules at various locations and the depth and width multiples are used to indicate the number of layers used for building the model, as shown in Table 1. However, YOLOv5x varies from other versions by using the PyCharm framework instead of Darknet. The architecture of YOLOv5 mainly consists of four blocks named input, backbone, neck, and head [44].

3.1.1. Input

The CT input is in Digital Imaging and Communications in Medicine (DICOM) format, it includes metadata, such as patient information, scanning acquisition time, hospital information, and pixel information. As a result, collected data should be pre-processed in order that the model can extract meaningful patterns from an input image. For this purpose, DICOM (.dcm) files should be converted into either.jpg or.png format using the pydicom library in the Python programming language [45]. The given CT slices are grayscale in nature, all input images in the dataset are displayed in a single window. In general, different types of hemorrhages are not visible in a single window because some specific features are highlighted in a particular window. Thus, to include this benefit in our proposed model, we used three different types of windows. The input CT image is composed of a series of CT numbers known as window width (WW) and window level (WL). Here, we obtained three different types of images through the windowing process, which are typically named bone window (WW: 1800, WL: 400), brain window (WW: 80, WL: 40), and subdural window (WW: 1800, WL: 400). The three images are stacked together to form a single three-channel RGB image as shown in Fig. 3, which has more information than any of the individual images. The images are resized to a specific size (256 × 256 × 3) in our proposed model for a simple training process. Initially, various data enhancement techniques, such as adaptive scaling algorithms or mosaic creation, are used in preprocessing CT input images. A mosaic means that the model will combine four randomly cropped input images to create a new input image. In mosaic creation, an increase in the number of input CT images enhances the robustness of the proposed model.

3.1.2. Backbone

The cross-stage partial (CSP) network is used as a backbone, to obtain quality features from the input while reducing the number of computations. An additional focus layer is used in slicing the input image into four 64 × 64 × 3 slices, which are then combined using a concat layer, yielding a result of 64 × 64 × 12. The combined images pass through a convolution layer with a kernel size of 32, and thus the output is 32 × 64 × 64. Then, batch normalization is performed, and the mish activation function (AF) is applied. The output of a mish AF is provided by equation (1), and the final expression of mish AF can be obtained by substituting the outputs of tanh and soft plus as shown in Eqn 2. After this process, C3 is used, which divides the feature maps from the previous layer into two parts. One part of the features pass through a dense block, which includes several convolutions, batch normalization, and mish activation, whereas the other part is a residual path. The outputs from these two paths are combined by using a concat layer. The

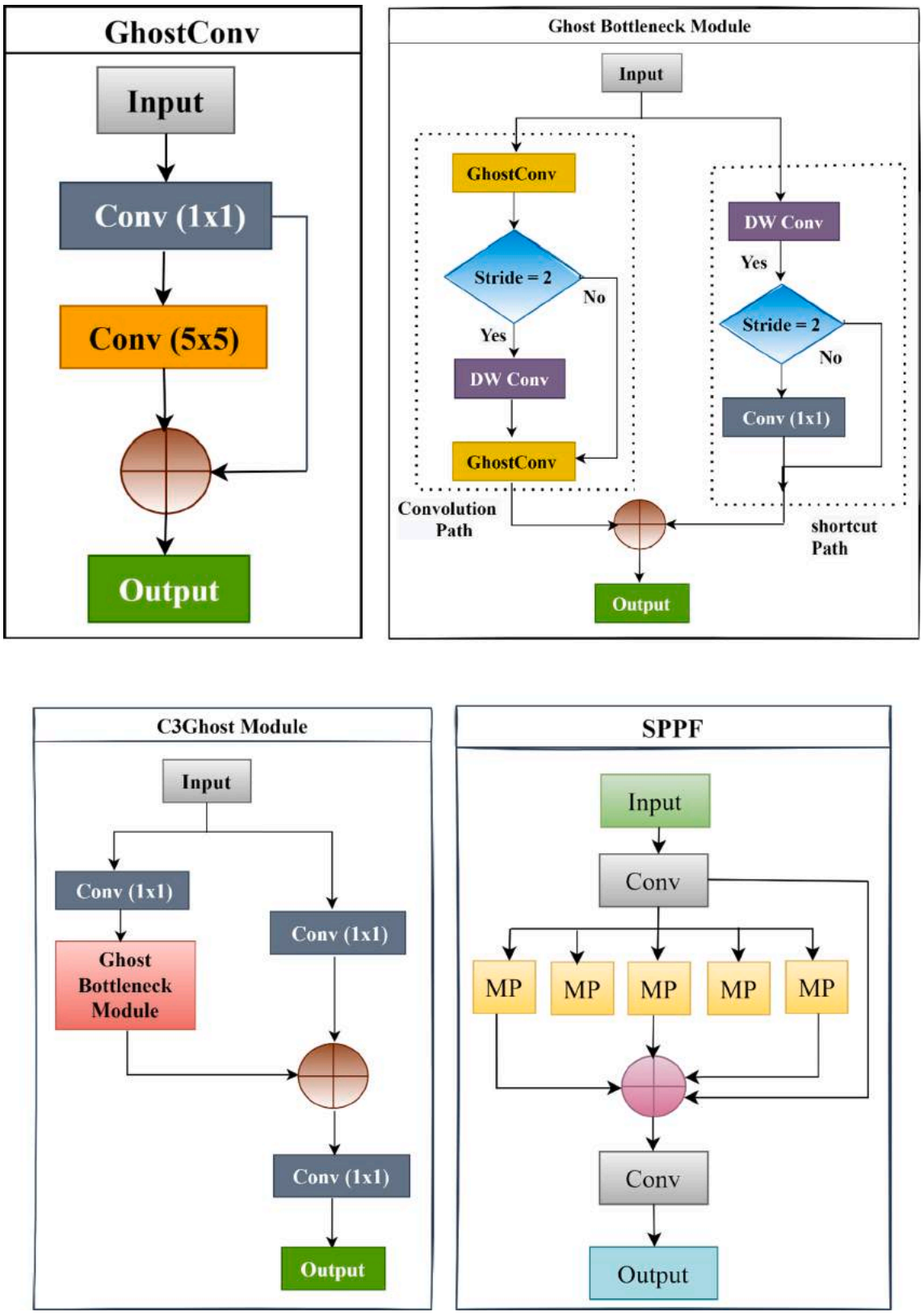

Fig. 5. Functional representation of (a) ghost convolution block, (b) ghost bottleneck module, (c) C3Ghost module, and (d) SPFF module.

main advantage of this process is that a model's learning capability can be improved by removing duplicate gradient information. The vanishing gradient problem is eliminated with this skip connection, and the number of computations is reduced as the feature maps are divided into two paths.

$$Mish(x) = x \times Tanh(softplus(x)) \tag{1}$$

$$Mish(x) = x\left\{\frac{\left[e^{ln(1+e^x)} - e^{-ln(1+e^x)}\right]}{e^{ln(1+e^x)} + e^{-ln(1+e^x)}}\right\} \tag{2}$$

Initially, we implemented the YOLOv5x model to enhance detection accuracy, but with this model, the number of computations and complexity increase substantially. Nevertheless, the traditional YOLOv5x has the highest accuracy for detecting mixed ICH, but computation power and time should be reduced by increasing FPS or reducing the number of parameters and floating-point operation per second (FLOPS). Moreover, a lightweight model can be used in real-time either by deploying it in cloud services or embedded devices with limited resources. To accomplish this, we proposed a novel YOLOv5x-GCB model that can produce similar results even with few resources by employing GC and C3Ghost modules in existing YOLOv5 , as shown in Fig. 4.

3.1.2.1. Ghost convolution (GC). In general, the output of a convolution layer yields several feature maps with redundant information because of the repetition of feature maps [46]. To prevent this situation, the proposed model includes a new concept known as GC. Compared with standard convolution, GC drastically reduces the complexity of the proposed model without sacrificing the number of feature maps. GC is a constant process regardless of the sizes of input feature maps, as shown in Fig. 5(a). Initially, a convolution layer with a kernel size of 1 × 1 is used to extract the intrinsic feature maps, after this, a 5 × 5 convolution is used to produce the remaining feature maps with less expensive linear operations. Finally, the outputs of the two layers are concatenated to form the final feature maps of GC.

3.1.2.2. C3Ghost. Based on the benefits of GC, we replaced the bottleneck CSP in the original YOLOv5 model with ghost bottleneck (GB) everywhere in the proposed model to obtain quality features from linear operations, which take a small amount of time to execute. GB includes two paths: the GC path and the shortcut path. A pair of GC modules are used in the GC path, then batch normalization and activations are used. The first GC acts as an expansion layer, increasing the number of channels. However, the second GC module reduces the number of channels, and a depth-wise convolution is used between the two modules to match the shortcut path. Finally, the output from these two paths is concatenated to produce the output of the GB module, as shown in Fig. 5(b). Finally, C3Ghost is designed with a GB module and three convolution layers that can significantly reduce the complexity of the proposed model, as shown in Fig. 5(c).

3.1.2.3. Spatial pyramid pooling faster (SPPF). The fully connected layers used in designing a model accept only fixed dimensional feature maps. Therefore, to generate fixed dimensions, we must resize our images every time, and this process decreases detection capability. To avoid this problem, SPPF is used as shown in Fig. 5(d), which generates feature maps with fixed dimensions regardless of the input size given to it by performing a max-pooling operation at five different levels [47].

3.1.3. Neck

In general, it is made up of a series of layers used in creating feature pyramids. The main objective of these feature pyramids is to recognize similar objects despite their different scales. To accomplish this task, a path aggregation network (PANET) is typically used as a neck, which performs three operations, namely, bottom-up path augmentation, adaptive feature pooling, and fully connected fusion [48]. Bottom-up-path augmentation is used in shortening an information path while improving feature pyramid values. To aggregate high and low-level feature maps, an adaptive feature pooling module is used. Finally, fully connected fusion is used in predicting objects at different scales. Thus, these feature pyramids play a vital role during the testing phase as data are unseen in the model previously [49].

3.1.4. Head or dense predictions

This is an output layer primarily used in making final predictions on the basis of input data. To make these predictions, the head uses data from the training phase, such as modified anchor box values, class probability values, and bounding box values. The head is made up of three convolutional layers with different input feature maps. So, the predicted output from these layers will have three different scales as well.

3.2. Evaluation metrics

In this paper, different metrics, such as recall, precision, F1 score, and mAP, with different threshold values were used in evaluating the detection of mixed ICH. The confusion matrix elements true positive, false positive, true negative, and false-negative were used for the evaluation of the proposed model. While implementing the YOLO model, we divided the input image of any size into S × S grids, and for each grid, B bounding boxes were created. A total of S × S × B boxes were generated, and the confidence score of each box can be calculated using equation (3),

$$CS_m^n = P_{m,n} \times IOU_{groundtruth}^{predict} \tag{3}$$

where, CS is the confidence score of the nth bounding box in an m^{th} grid cell, and $P_{m,\,n}$ is the probability of hemorrhage, and its value ranges between 0 and 1. The amount of intersection between the ground truth and predicted bounding boxes is denoted by IOU.

Recall is used in calculating the proportion of correctly predicted positive ICH cases in all positive samples and is given by equation (4). The graph of recall is obtained by considering the recall value at different confidence levels.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4}$$

Precision is used in determining whether the proposed model's prediction of various types of ICH is comparable to the actual type of ICH. The precision curve is plotted by considering the precision value at various confidence levels. The precision is calculated by using equation (5),

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

F1-Score is calculated by taking the harmonic mean of recall and precision, and its value becomes unity when the precision and recall values are 1. It is expressed mathematically according to equation (6).

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{6}$$

Precision and Recall Curve is a plot drawn by considering the precision on the y-axis and recall on the x-axis. As this curve is a tradeoff between precision and recall, the proposed model produces good results if the top right-hand side of the curve has a high value.

Mean Average Precision (mAP) is an important metric in object detection and is calculated by taking the average of average precision (AP) values for each individual class. The mathematical representation of mAP is given by following equation (7), and it is calculated by fixing the threshold value.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} APi \tag{7}$$

where, N is the number of classes, and the AP is used to represent the area under the precision and recall curve as shown in Eqn 8. where p is the precision rate and r is the recall rate.

$$AP = \int_0^1 p(r)dr \tag{8}$$

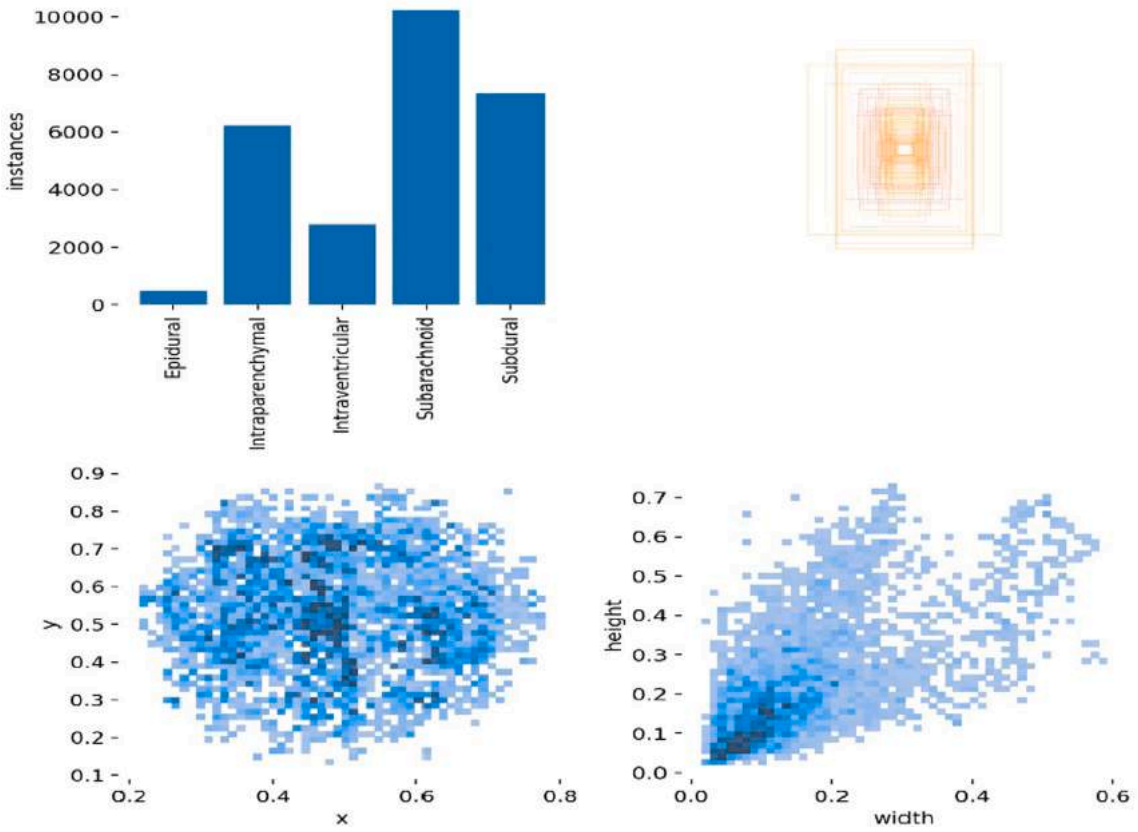

Fig. 6. (a) Bar graph that shows the count of different types of hemorrhages in the training dataset, (b) Bounding box distribution, (c) Statistical distribution of bounding box position, and (d) Statistical distribution of bounding box size.

4. Results

4.1. Description of datasets

In the evaluation of the performance of the proposed model, CT image data was collected primarily from two repositories. Two separate publicly available datasets were used in training and testing the proposed model's performance. The first dataset was a brain hemorrhage extended (BHX) dataset, which contained information on 491 patients [50]. Of these patients, 205 tested positive, and the remaining were labeled as normal. Only information from positive patients was used for training and validation. Thus, a total of 21,132 slices were extracted from 205 patients, which includes all the different types of hemorrhages, and a few slices had mixed hemorrhages. Here, the annotations were created by expert radiologists, and the total number of different types of hemorrhages, as well as their bounding box information, is shown in Fig. 6. Fig. 6(a) is a bar graph that shows the number of bounding boxes for each individual class, whereas Fig. 6(b) shows all of the bounding boxes in the dataset. Fig. 6(c) and 6(d) depict the statistical distribution of the bounding box position and size. These two figures are derived from the histogram plot and are primarily used to visualize the distribution of various data points in the employed dataset.

The second dataset was a segmentation dataset consisting of 2500 slices from 75 patients [23]. Only 318 of the slices had ICH, whereas the rest were normal. In these 318 images, masks that showed the exact locations of the hemorrhages were provided in a separate set of images. Initially, the contours of the masks were extracted, and then a rectangle box was created to generate the ground truth. Here, LabelImg software was used in creating the ground truth labels in YOLO-accepted format [51]. Finally, the images were used in testing the robustness of the proposed model in a real-world scenario by overlapping the original ground truth box with the predicted box.

Table 2
Statistical analysis of the CQ500 dataset.

Finding		EDH	IPH	IVH	SAH	SDH	Overall
Reader 1, 2	Agreement %	97.35	91.24	96.13	93.08	87.98	89.00
	Cohen's K	0.505	0.786	0.704	0.677	0.485	0.777
Reader 1,2	Agreement %	98.37	90.63	97.15	90.84	93.08	90.84
	Cohen's K	0.725	0.765	0.735	60.58	0.600	0.808
Reader 1, 2	Agreement %	98.17	90.84	95.72	90.84	90.02	88.39
	Cohen's K	0.599	0.771	0.655	0.636	0.562	0.764
Reader 1, 2, 3	Fleiss' κ	0.614	0.774	0.696	0.638	0.541	0.782
Bounding Box Count		587	7244	2432	9951	7494	27,708
Mean		0.026	0.330	0.111	0.454	0.342	–
Standard Deviation		0.161	0.470	0.314	0.497	0.474	–
Variance		0.026	0.221	0.098	0.247	0.225	–

EDH: Epidural hemorrhage, IPH: Intraparenchymal hemorrhage, IVH: Intraventricular hemorrhage, SAH: Subarachnoid hemorrhage, SDH: Subdural hemorrhage.

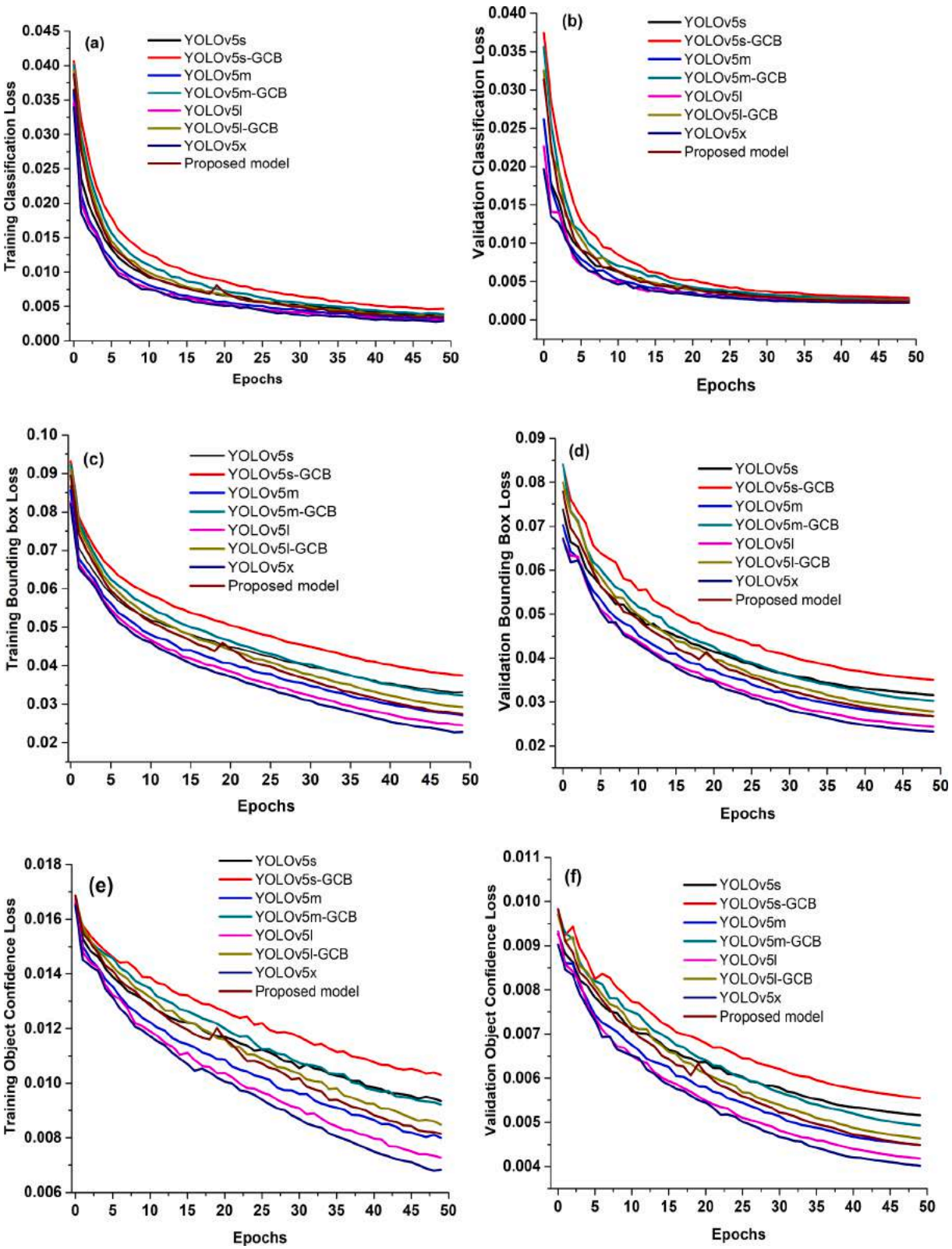

Fig. 7. Plots of training and validation loss with respect to the number of epochs for all the variants YOLOv5, including the proposed model: (a) training classification loss, (b) validation classification loss, (c) training bounding box loss, (d) validation bounding box loss, (e) training object confidence loss, and (f) validation object confidence loss.

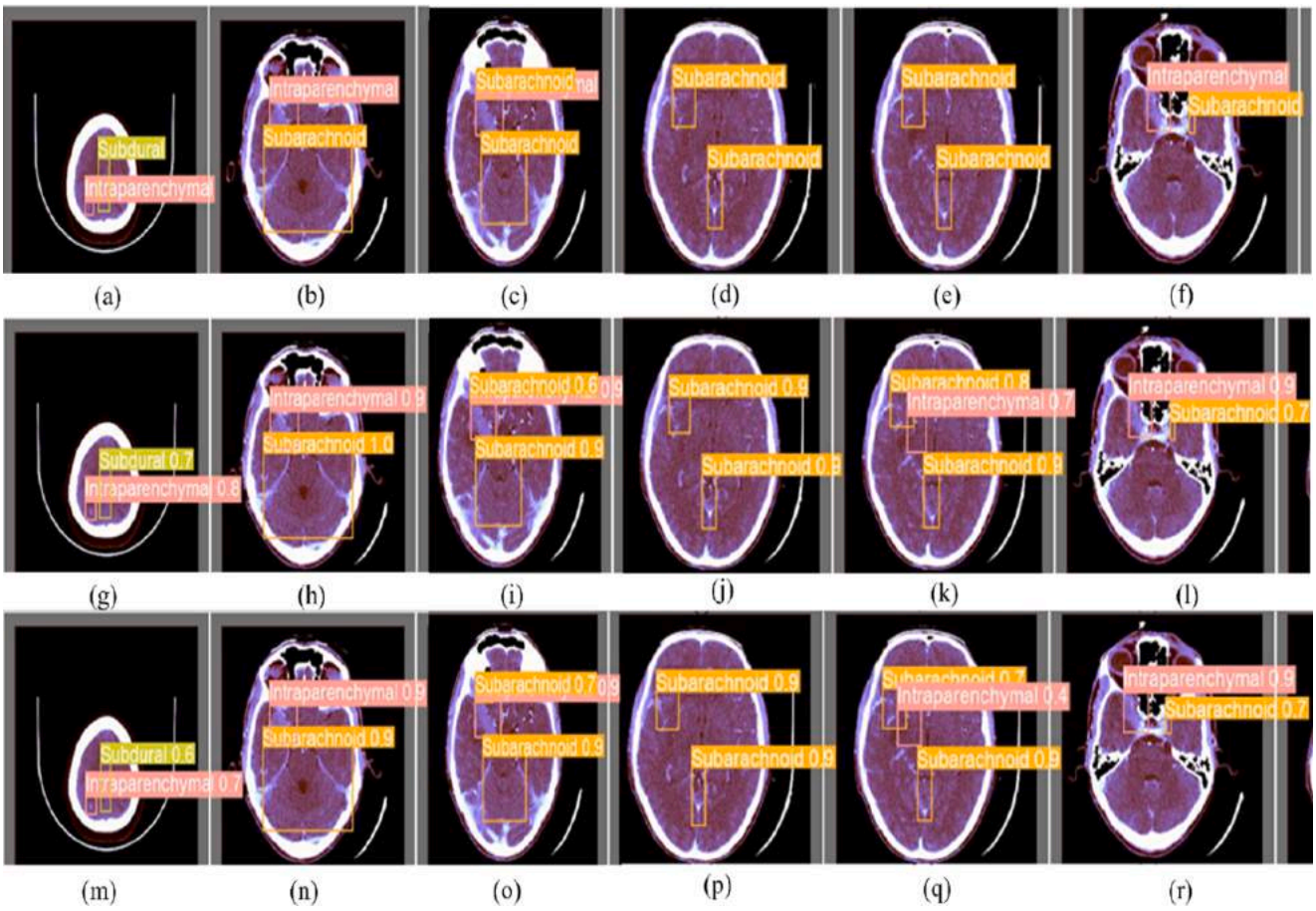

Fig. 8. A batch of 6 images is used in the validation of the proposed model: (a–f) ground truth images of the ICH consisting of a bounding box along with its label name; (g–l) predicted images by YOLOv5x that shows the bounding box and confidence score value; and (m–r) predicted images by the proposed YOLOv5x-GCB model.

4.1.1. Statistical analysis

We calculated the inter-rater reliability between every-two radiologists using Cohen's kappa (K) and the percentage of agreement [35]. In addition, the reliability between all three readers was measured by Fleiss' kappa(κ) coefficient. The values of Cohen's kappa are interpreted as below: 0.91–1.0 as excellent, 0.81–0.90 as very good, 0.61–0.80 as good, 0.41–0.60 as moderate, and less than or equal to 0.3 as poor. The sample sizes of each hemorrhage and all the statistical measures such as mean, variance, and standard distribution values were calculated in the statistics library in python programming and are tabulated in Table 2.

4.2. Model training and loss function

Initially, model training was stated with pretrained COCO weights (mostly used in object detection). However, these weights were unsuitable for our data [52]. Thus, during the training process, the weight values were adjusted for every epoch by using equation (9), where W_{new} is the new weight values after the training process, W_{old} is the old weight values before the training, η is the learning rate, and L is the loss value. The training process was completed whenever the loss value converged to a minimum, and the loss value in YOLOv5 was the sum of three individual losses, as shown in equation (10).

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W_{old}} \tag{9}$$

$$LOSS(YOLOv5x - GCB) = LOSS_{BoundingBox} + LOSS_{Classification} + LOSS_{confidence} \tag{10}$$

The bounding box loss of a specific box named 'p' is caused by either the dimensions of the box (w_p, h_p) or the position of the box (x_p, y_p). Hence, this loss is a combination of two individual terms, as shown in equation (11). The coordinates (x_p, y_p, w_p, h_p) represent the actual dimensions of the box, and ($\hat{x}_P$, $\hat{y}_P$, $\widehat{W_P}$, $\widehat{h_P}$) represents the predicted one.

$$\begin{aligned} LOSS_{BoundingBox} = \phi_{Position} \Bigg\{ \sum_{m=0}^{s^2} \sum_{n=0}^{B} \Omega_{mn}^{ICH} \left[(x_P - \hat{x}_P)^2 \right. \\ \left. + (y_P - \hat{y}_P)^2 \right] \Bigg\} + \phi_{Position} \Bigg\{ \sum_{m=0}^{s^2} \\ \times \sum_{n=0}^{B} \Omega_{mn}^{ICH} \left[\left(\sqrt{W_P} - \sqrt{\widehat{W}_P} \right)^2 + \left(\sqrt{h_P} - \sqrt{\widehat{h}_P} \right)^2 \right] \Bigg\} \end{aligned} \tag{11}$$

where $\Phi_{Position}$ is used to penalize the false position, and Ω_{mn}^{ICH} specifies that the n^{th} bounding box in an m^{th} grid cell is responsible for prediction and its value will become "1" if hemorrhage is present and otherwise "0". Here, a square root was used for the height and width values to penalize the errors in large bounding boxes rather than the errors in small boxes. The classification loss was similar to the general binary classification loss, as shown in equation (12).

$$LOSS_{Classification} = \sum_{m=0}^{s^2} \Omega_m^{ICH} \sum_{C \in class} \left[P_m(c) - P_m(\bar{c}) \right]^2 \tag{12}$$

where the value of Ω_m^{ICH} is "1" if the m^{th} grid cell contains hemorrhage and otherwise "0". The P_m(c) value was used to indicate the ground truth

Table 3
The predicted confidence score for a set of 6 input images during the validation stage by both the YOLOv5x and the proposed models.

Figure Number	Label Name(s)	Predicted confidence Score by YOLOv5x	Predicted confidence Score by Proposed Model
8(a)	Subdural, Intraparenchymal	0.7, 0.8	0.6, 0.7
8(b)	Intraparenchymal, Subarachnoid	0.9, 1.0	0.9, 0.9
8(c)	Intraparenchymal, Subarachnoid, Subarachnoid	0.9, 0.6,0.9	0.9, 0.7,0.9
8(d)	Subarachnoid, Subarachnoid	0.9, 0.9	0.9, 0.9
8(e)	Subarachnoid, Subarachnoid	0.8, 0.9 (FP: Intraparenchymal-0.7)	0.7, 0.9 (FP: Intraparenchymal-0.4)
8(f)	Intraparenchymal, Subarachnoid	0.9, 0.7	0.9, 0.7

FP: False Positive.

output for a specific class, whereas $P_m(\bar{c})$ was used for the predicted output. Finally, equation (13) represents the confidence loss,

$$LOSS_{Confidence} = \sum_{m=0}^{s^2} \sum_{n=0}^{B} \Omega_{mn}^{ICH} \left[(cs_m - \overline{cs_m})^2 \right] + \lambda_{no-ICH} \sum_{m=0}^{s^2} \times \sum_{n=0}^{B} \Omega_{mn}^{no-ICH} \left[(cs_m - \overline{cs_m})^2 \right] \quad (13)$$

where, the value of Ω_{mn}^{no-ICH} is opposite to Ω_{mn}^{ICH}, while λno-ICH is similar to $\Phi_{Position}$. The term cs_m represents the actual confidence score, and $\bar{c}s_m$ represents the predicted confidence score. Once the loss value is calculated for a single epoch, a process known as optimization is used to converge this loss to a minimum point. Numerous types of optimizers are available, and out of them, stochastic gradient descent (SGD) was selected as an optimizer because of its advantages over others. SGD follows a zigzag path to reach a minimum point, a parameter called momentum was used, which nullified that effect and accelerated the process of reaching the global minimum point. Once the proposed model was properly built, it was trained with labeled data for predictions on test data. In training and validation, three individual losses: bounding box loss, classification loss, and confidence loss were generated by equations (11), 12, and 13. Fig. 7 shows the training and validation loss curves for the proposed and other versions of YOLOv5 models as a

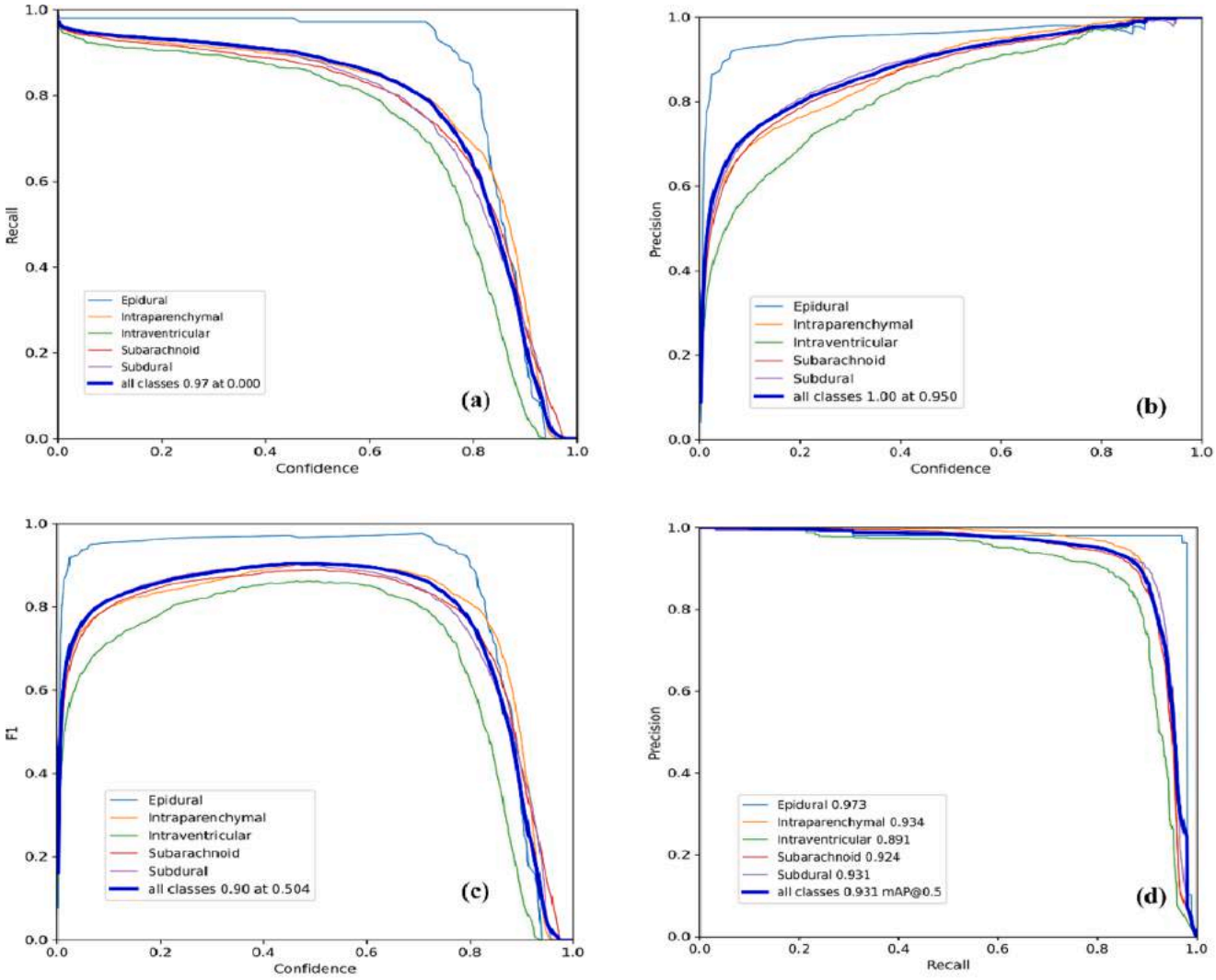

Fig. 9. Confidence score with respect to performance metrics of the proposed model for all different types of ICH: (a) Recall, (b) Precision, (c) F1 score, and (d) Precision vs Recall, which is used to determine the value of mAP at 0.5 threshold value.

Table 4
Performance metrics of existing (YOLOv4) and proposed (YOLOv5x-GCB) model with repect to each hemorrhage class.

Name of the Class	Precision		Recall		F1-Score		mAP@0.5		mAP@0.5:0.95	
	Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed	Existing	Proposed
EPH	1.00	**0.964**	1.00	**0.971**	1.00	**0.967**	1.00	**0.973**	–	**0.685**
IPH	0.97	**0.929**	0.86	**0.881**	0.91	**0.904**	0.85	**0.934**	–	**0.647**
IVH	0.68	**0.877**	0.65	**0.847**	0.67	**0.861**	0.63	**0.891**	–	**0.516**
SAH	0.97	**0.911**	0.74	**0.867**	0.84	**0.888**	0.71	**0.924**	–	**0.643**
SDH	0.97	**0.921**	0.80	**0.878**	0.88	**0.898**	0.79	**0.931**	–	**0.629**
Overall	0.92	**0.921**	0.81	**0.889**	0.86	**0.900**	0.796	**0.931**	–	**0.624**

Exiting model (YOLOv4) and Proposed model (YOLOv5x-GCB).

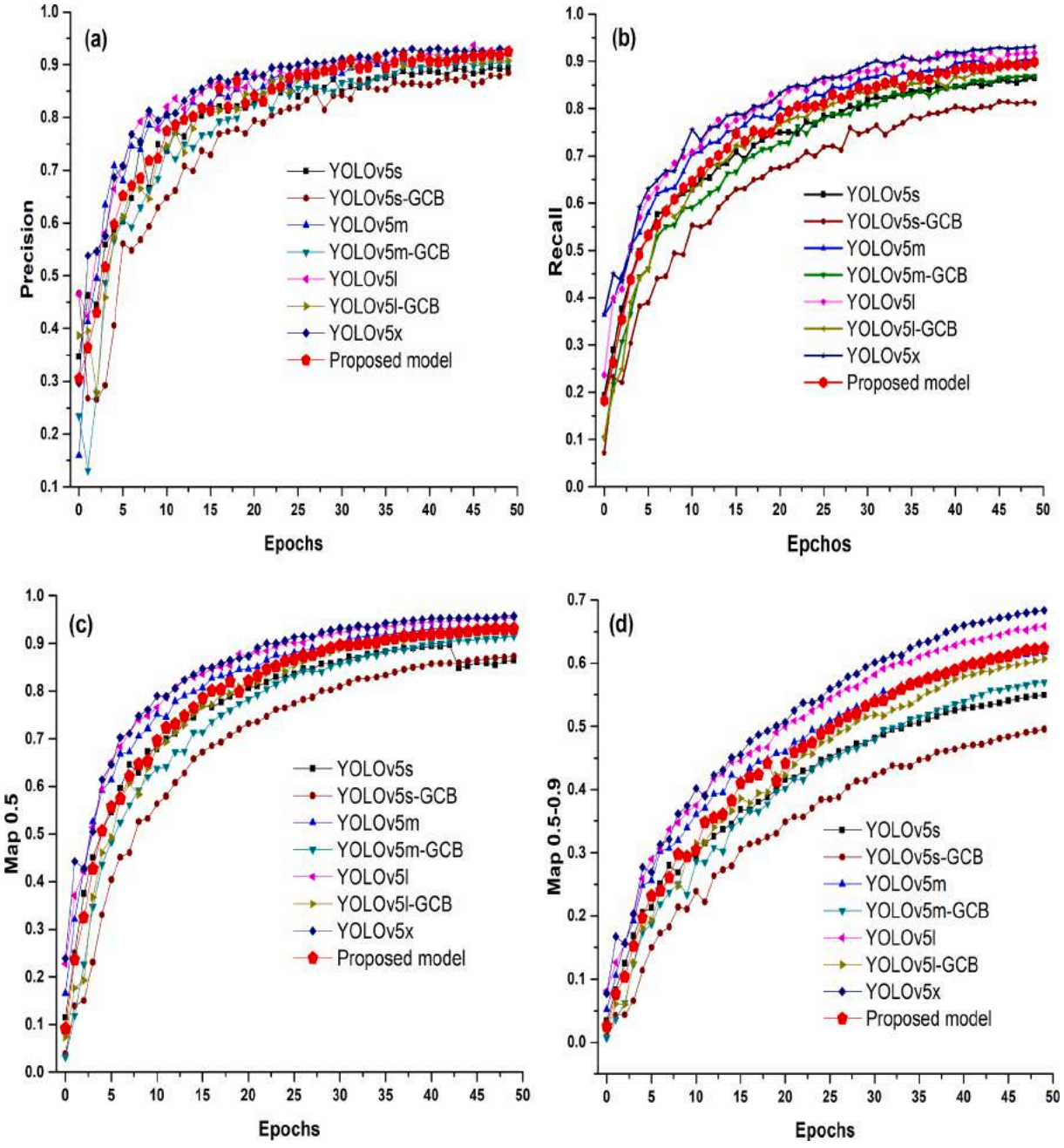

Fig. 10. Comparison of the performance metrics of the proposed model (YOLOv5x-GCB) and existing models with respect to the number of epochs: (a) precision curve, (b) recall curve, (c) mAP curve with 0.5 as a threshold, and (d) mAP curve with 0.5–0.9 as a threshold value.

function of the number of epochs. These plots were produced by considering the loss values at each epoch, and they demonstrated the proposed model fine convergence.

4.3. Experimental setup

In the training of the proposed model, the total available data was divided into an 80:20 ratio for training and validation, yielding 16,905 and 4227 training and validation images. During the training process, a batch size of 16 was used, resulting in a total of 1057 (16905/16) iterations for each epoch. In validation, a batch size of 32 was used, resulting in a total of 133 (4227/32) iterations for each epoch. The proposed model set the momentum value at 0.937, with a weight decay factor of 0.0005. In general, the value of η was determined by either a static rule or a dynamic learning rate rule. In the static method, the η value is selected in such a way that it should not be extremely high or low. In the dynamic rule, its value varies according to a specific condition. Here, the dynamic learning rate was used, where the initial learning rate was set at 0.01, and its value became 0.1 after 50 epochs. The threshold value for IOU was set at 0.5. Finally, the proposed model was executed in the cloud-based Google Pro platform using a Tesla P100 GPU and PyCharm framework.

4.4. Bounding box detection

In this study, a GC-based YOLO (YOLOv5x-GCB) architecture was used to classify mixed hemorrhages in a CT slice. To achieve this task, the proposed model used the bounding box technique in selecting a hemorrhage region instead of a normal segmentation mask. The proposed architecture was trained with the BHX dataset for the evaluation of training and validation accuracy. Finally, the performance of the proposed model was measured using four metrics, namely, precision, recall, F1 score, and mAP, in terms of confidence score with respect to 50 epochs. During training, a batch size of 16 was used at a time, allowing us to visualize the number of 16 images as a batch for predicting the mixed ICH. Fig. 8 depicts a sample of 6 images with mixed ICH cases for the ground truth (Fig. 8(a-f)), predicted output with standard YOLOv5x (Fig. 8(g-l)), and predicted output with the proposed model (Fig. 8(m-r)). Table 3 shows the confidence score values for images predicted by YOLOv5x and the proposed model. According to the values in , Table 3 the proposed model produced results similar to those of the standard YOLOv5x.

4.5. Evaluation results on BHX dataset:

The proposed algorithm is trained with different types of hemorrhages. Fig. 9 depicts the proposed model performance metrics as the functions of the confidence score for each ICH class separately. Fig. 9(a), (b), and (c) show that epidural hemorrhage had the highest precision, recall, and F1 score among all classes. The PR curves for all ICH classes are shown in Fig. 9(d), which indicates that the proposed model had the highest predicted precision at a reasonable recall. Despite that the number of input epidural hemorrhage samples was few, it had the highest precision, whereas intraventricular hemorrhages had the lowest precision. The reason was that EDH is straightforward in detection, whereas the IVH was surrounded by other tissues, which made their detection difficult. The metric values of the proposed and existing (YOLOv4) models for each of the five individual classes are shown in Table 4. The proposed model significantly improved recall, F1 score, and mAP@0.5 for all types of hemorrhages.

The overall performance for all types of hemorrhages indicated that the proposed model precision was 0.921, which was similar to the values obtained with the existing model, and the metrics (recall, F1 score, and mAP@0.5) had values of 0.889, 0.900, and 0.931, respectively, which significantly improved relative to those in the existing model. The mAP@0.5–0.9 threshold determined for the proposed model is presented in Table 4. The bold data in Table 4 represents the proposed model performance.

Table 5
Comparison of precision, recall, F1 score, and mAP values for conventional models, YOLO models, and proposed model.

Name of the Model	Precision	Recall	F1-Score	mAP@0.5	mAP@0.5:0.95
3D-CNN + YOLO: HR [41]	0.619	–	0.747	–	–
3D-CNN + YOLO: LR [41]	0.672	–	0.776	–	–
YOLOv2 + single label [42]	0.609	–	0.695	–	–
YOLOv2 + double labels [42]	0.627	–	0.611	–	–
YOLOv2 + CSF filtering [42]	0.797	–	0.727	–	–
SSD(512)-FE [40]	0.797	–	0.845	–	–
2D Faster RCNN [38]	0.897	–	0.908	–	–
Faster R-CNN [39]	0.857	0.844	0.85	–	–
R-FCN [39]	0.905	0.826	0.864	–	–
YOLOv4 -ST 43]	0.938	0.918	0.928	0.906	–
YOLOv4 -TT [43]	**0.920**	**0.810**	**0.860**	**0.796**	**–**
YOLOv5s -TT	0.895	0.863	0.880	0.909	0.550
YOLOv5s- GCB -TT	0.885	0.841	0.850	0.873	0.510
YOLOv5m -TT	0.910	0.900	0.910	0.920	0.617
YOLOv5m – GCB -TT	0.905	0.869	0.890	0.915	0.572
YOLOv5l -TT	0.929	0.918	0.920	0.934	0.638
YOLOv5l –GCB -TT	0.909	0.892	0.900	0.929	0.608
YOLOv5x -TT	0.928	0.931	0.930	0.948	0.663
Proposed Model (YOLOv5x-GCB) -TT	**0.921**	**0.889**	**0.900**	**0.931**	**0.624**

* ST – Separate Training, TT- Together Training, HR – High Resolution, LR-Low Resolution, and FE –Feature Enhancement.

Fig. 10 depicts a comparison of the proposed model's performance metrics with the other versions of YOLOv5. With respect to the number of epochs, all evaluation metrics, such as precision (Fig. 10a), recall (Fig. 10b), mAP@0.5 (Fig. 10c), and mAP@0.5–0.9 (Fig. 10d) gradually increased for all models. In these graphs, the proposed model provided an average value in all metrics with respect to the number of epochs. Table 5 shows the comparison of the proposed model with the other state-of-the-art methods. Here the results are mainly compared with the existing model YOLOv4. The recall, F1 scores, and mAP values of the proposed model improved, and precision was comparable to that of the existing method, indicating that the proposed model outperformed YOLOv4.

4.6. Detection results on test dataset

After the completion of the training process, the proposed model was tested with unseen data. Here, the phrase unseen data refers to new separate data, that is, it is not a part of either training or validation. For this purpose, we used the second dataset. Fig. 11 shows a sample of the predicted images during the test phase. The blue box represents the ground truth bounding box, whereas the other color represents the predicted box with the proposed model.

4.7. Ablation study on the proposed model

Table 6 shows the results of ablation experiments to confirm the effect of mosaic data augmentation on the detection performance of ICH.

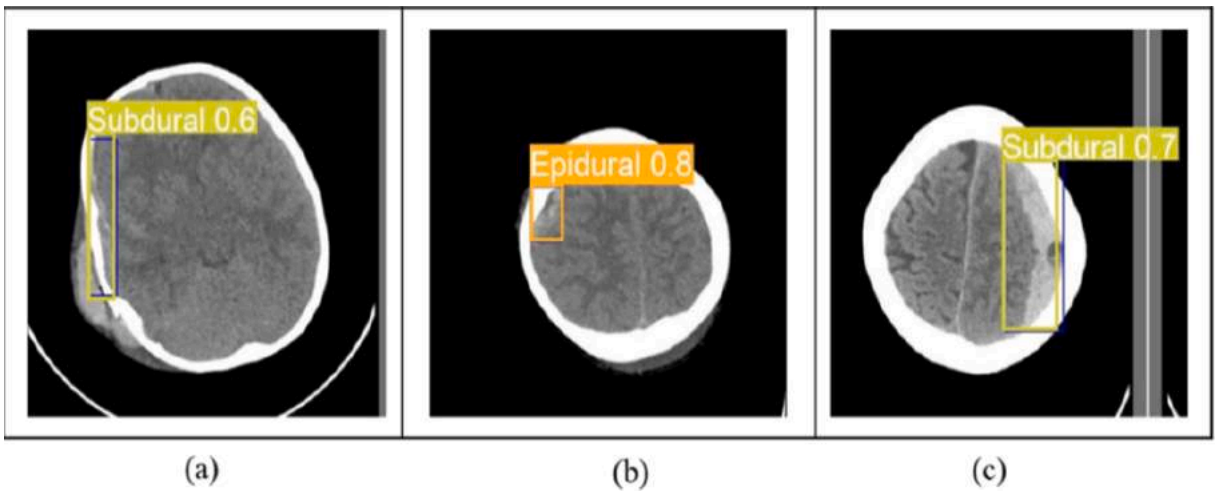

Fig. 11. Predicted images during the testing phase of the proposed model for various types of ICH: (a) subdural type with a confidence score of 0.6; (b) epidural type hemorrhage with a confidence score value of 0.8; and (c) subdural hemorrhage with a confidence score of 0.7. The blue color represents the ground truth bounding box, and the yellow and orange colors represent the predicted bounding boxes.

Table 6
The performance of the proposed model for detecting ICH using mosaic data ablation.

Name of the Model	Use of Mosaic	Precision	Recall	F1 score	mAP@0.5	mAP@0.5:0.95
YOLOv5s	×	0.646	0.541	0.588	0.585	0.262
YOLOv5m	×	0.711	0.596	0.648	0.633	0.291
YOLOv5l	×	0.757	0.647	0.697	0.693	0.336
YOLOv5x	×	0.792	0.673	0.727	0.721	0.349
Proposed Model	×	0.788	0.652	0.713	0.699	0.323
Proposed Model (YOLOv5x-GCB) -TT	√	**0.921**	**0.889**	**0.900**	**0.931**	**0.624**

× - Model trained without Mosaic √ - Model trained with Mosaic.

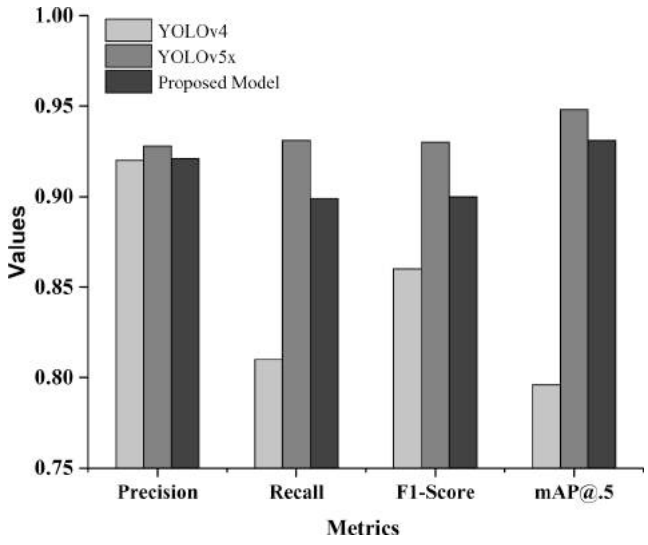

Fig. 12. A bar graph that shows the metrics comparison of the proposed model (YOLOv5x-GCB), existing YOLOv4, and standard YOLOv5x.

Experiment results show that when the Mosaic concept is used, the detection results of the proposed model were improved significantly.

5. Discussions

This research provides a comprehensive analysis of the localization of mixed ICH by using the YOLOv5x-GCB architecture with limited resources. The classification of multiple hemorrhages in real-time by using conventional methods and existing CNN models is not possible because these algorithms are effective if and only if a patient has only one type of hemorrhage. As a result, an object detection algorithm was used in the proposed model to classify mixed hemorrhages in the given CT image. In addition, the proposed model takes less convergence time and uses fewer resources. Table 5 indicates that the proposed model outperformed existing models for all classes in terms of performance metrics. The existing (YOLOv4) and proposed methods in Table 5 employed a CT image dataset, whereas the remaining methods in Table 5 used the MRI dataset for model training. Fig. 12 depicts a comparison of the evaluation metrics of YOLOv4, standard YOLOv5x, and the proposed model. As shown in Fig. 12, the proposed model evaluation parameters were similar to the YOLOv5 model and better than the existing YOLOv4 model. These findings indicated that the proposed model outperformed existing models in predicting multiple hemorrhages in a single slice. However, the disadvantage of standard YOLOv5x is that it requires a large number of resources and time to complete the training process. To address this, we proposed the YOLOv5x-GCB model, which produces nearly identical detection results with optimized resources, as shown in Table 7.

Initially, we implemented the YOLOv5s model, and its performance is compared with YOLOv4 as shown in Table 5. Here, the recall value increased by 5.3 % compared with that value in YOLOv4, and the mAP and F1 scores increased by 11 % and 2 %, respectively, but the precision decreased by 2.5 % (Table 5). To improve the detection capability furthermore, we implemented other versions of YOLOv5 by increasing the network depth, such as YOLOv5m, YOLOv5l, and YOLOv5x. Finally, using the existing YOLOV5x, we achieved 0.928 precision, 0.931 recall, 0.93 F1 score, and 0.948 mAP@0.5. With these results, we confirmed that YOLOV5x produced significantly better results than the existing models. However, increasing the number of layers for better performance increases the values of other parameters, such as gradient,

Table 7
Comparison of hardware requirements and the number of calculations used in conventional and proposed models and parameters used in determining the best model.

Model YOLO	Layers	Parameters (Million)	Gradients (Millions)	Flops	Model Weight (MB)	Frames/second (FPS)	Iterations/ second (IPS)	Memory GB (GPU)	Avg. time/ epoch (Min)	Total Execution Time (Hours)
v5s	270	7.033	7.033	15.9	14.3	187	11.65	0.751	1.30	1.490
v5s- GCB	453	3.695	3.695	8.10	7.70	165	10.27	0.629	1.42	1.540
v5m	369	20.88	20.88	48.1	42.1	94	5.86	1.400	3.00	2.851
v5m - GCB	695	8.542	8.542	18.4	17.5	123	7.67	1.170	2.15	2.140
v5l	468	46.15	46.15	108.0	92.7	60	3.77	2.240	4.41	4.320
v5l -GCB	937	15.61	15.61	33.3	31.8	95	5.92	1.770	2.56	2.650
v5x	**567**	**86.24**	**86.24**	**204.3**	**173.1**	**33**	**2.10**	**3.780**	**8.22**	**7.752**
Proposed Model	**1179**	**25.08**	**25.08**	**53.3**	**50.9**	**59**	**3.68**	**2.350**	**4.47**	**3.320**

GFLOP, total weight size, GPU-memory required, and execution time, as shown in Table 7. Although, the primary goal of this research is to develop a model that can be able to detect mixed hemorrhages with limited resources.

To achieve this objective, we used the ghost convolution module in standard YOLOv5. With this process, although the number of layers increased to a large number, the computation time, FLOPs, and memory were drastically reduced. The GC process was initially implemented in YOLOv5s, and then the number of layers was increased from 270 to 453. However, the number of gradients, FLOPs, and model weights was reduced by nearly half. However, as the original YOLOv5s itself had fewer layers, we were unable to find a positive change in FPS, the number of iterations per second, and the total average execution time of YOLOv5s-GCB. Therefore, to examine these changes in the right direction, we implemented this ghost concept in m, l, and x as YOLOv5m-GCB, YOLOv5l-GCB, and proposed YOLOv5x-GCB models. Finally, with the proposed model, we achieved a lightweight model with similar performance results to the original YOLOv5x model.

The main advantages of the proposed model are as follows:

(1) It can detect the mixed ICH along with their localization.
(2) It requires less memory, allowing us to deploy the proposed model in either cloud or in embedded devices for real-time clinical diagnosis.
(3) It creates a bounding box around a hemorrhage, thus improving segmentation accuracy, and it is simpler than pixel-wise semantic segmentation. Moreover, a radiologist may obtain additional information about the injury location.
(4) The total time required to execute the proposed model is reduced as it has a higher FPS and IPS.
(5) The number of FLOPs required to execute the proposed model is extremely low even though the number of layers increases.

The main limitations of the proposed model are as follows: (1) The addition of a ghost module in the standard YOLOv5x leads to the degradation of performance metrics. (2) The dataset used in executing the proposed model is imbalanced. (3) The bounding box for each input image is selected manually in the preparation of training data.

These limitations can be addressed by improving the proposed model by incorporating the squeeze and excitation (SE-module) in the backbone, or the concept of ensemble learning can be used to promote the detection capability even in the presence of a ghost module. The data imbalance problem can be eliminated by assigning the highest weight value to a minority class and the lowest weight to a majority class label. Multiple hemorrhages commonly occur in real-time clinical diagnosis. Therefore, detection and classification mechanisms deployed "on the fly" in a CT scanner would be extremely valuable while performing the clinical diagnosis. In the clinical implementation stage, the proposed algorithm generally works behind the scenes to optimize the radiologist's read time for studies with ICH. In general, radiologists analyze CT scans at the top of the reading list. As a result, a ruled-based engine can be used to sort the reading list by prioritizing emergency examinations. Finally, a data pipeline system can transfer CT studies to the computing server that contains the proposed algorithm, allowing for real-time implementation of the proposed algorithm. The algorithm generates a binary output after processing a specific CT study (either positive or negative ICH). If the results were positive, the priority of the study was raised to "emergency," and the radiologist reading list was immediately updated. On the other hand, if the findings were non-emergency, the study's priority remained unchanged.

6. Conclusions

The primary goal of this research work is to create a model that can detect the presence of mixed hemorrhages in a CT image by using a lightweight model with limited resources. To achieve this objective, a standard YOLOv5x model was used, and the accuracy of the detection results were ensured by predicting the mixed hemorrhages in a single slice. However, the limitation of the standard YOLOv5x is that it requires large memory and takes a longer time to make predictions. To overcome this limitation, a novel YOLOv5x-GCB model was proposed, which employs the concepts of GB and GC. The main benefit of incorporating these modules into the proposed model is that the memory required to store the weights of the proposed model is reduced to nearly-one-third that of the existing model, allowing us to deploy it in real-time clinical diagnosis by using the cloud services to optimize the radiologists read time. Moreover, the speed increases two times, and the number of computations are reduced to one-quarter of the time by producing nearly identical results. Furthermore, the significance of semantic segmentation is greatly enhanced as a result of the proposed bounding box concept.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Lakshmi Prasanna Kothala: Methodology, Formal analysis, Investigation. **Prathiba Jonnala:** Writing – review & editing. **Sitaramanjaneya Reddy Guntur:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data used for this work is brain hemorrhage extended (BHX) dataset, which is publicly available

References

[1] S. Parikh, K. Marcella, R.K. Narayan, Traumatic brain injury, Int Anesthesiol Clin 45 (3) (2007) 119–135.
[2] J.J. Heit, M. Iv, M. Wintermark, Imaging of intracranial hemorrhage, J Stroke 19 (1) (2017) 11–27.
[3] P.D. Chang, E. Kuoy, J. Grinband, B.D. Weinberg, M. Thompson, R. Homo, J. Chen, et al., Hybrid 3D/2D convolutional neural network for hemorrhage evaluation on head CT, J. Neuroradiol 39 (9) (2018) 1609–1616.
[4] U. Balasooriya, M.S. Perera, Intelligent brain hemorrhages diagnosis using artificial neural networks, Business Engineering and Industrial Applications Colloquium (BEIAC) (2012) 128–133.
[5] T. Chan, Computer aided detection of small acute intracranial hemorrhage on computer tomography of brain, Comput Med Imaging Graph 31 (4–5) (2007) 285–298.
[6] C.J. Asch, M.J. Luitse, G.J. Rinkel, I. Tweel, A. Algra, 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 9 (2) (2010) 167–176.
[7] J.C. Hemphil, S.M. Greenberg, C.S. Anderson, K. Becker, B.R. Bendok, M. Cushman, et al., Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the american heart association/ American stroke association, Stroke 46 (7) (2015) 2032–2060.
[8] F. Piccialli, V.D. Sommaa, F. Giampaoloa, S. Cuomoa, G. Fortino, A survey on deep learning in medicine: why, how and when? Inf. Fusion 66 (2021) 111–137.
[9] T. Magadza, S. Viriri, Deep learning for brain tumor segmentation: a survey of state-of-the-art, J. Imaging 7 (2) (2021) 19.
[10] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.W. M. Laak, B. Ginneken, C.I. Sánchez, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[11] X.B. Nguyen, G.S. Lee, S.H. Kim, H.J. Yang, Self-supervised learning based on spatial awareness for medical image analysis, IEEE Access 8 (2020) 162973–162981.
[12] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit (2014) 580–587.
[13] R. Girshick, R.-C.-N.-N. Fast, IEEE Int. Conf. Comput. Vis (2015) 1440–1448.
[14] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, 28th Int, Conf. Neural Inf. Process 28 (2015) 291–299.
[15] K. He, G. Georgia, P. Dollár, R. Girshick, Mask r-cnn,, IEEE Int. Conf. Comput. Vis (2017) 2980–2988.
[16] W. Liu, A. Dragomir, E. Dumitru, S. Christian, R. Scott, Y.F. Cheng, C.B., S.S.D. Alexander, Single shot multibox detector, European Conference on Computer Vision, (2016) 21-37.
[17] R. Joseph, D. Santosh, R. Girshick, F. Ali, You only look once: Unified, real-time object detection, IEEE conference on computer vision and pattern recognition. (2016) 779-788.
[18] R. Joseph, F. Ali, YOLO9000: better, faster, stronger, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit, (2017) 6517-6525.
[19] R. Joseph, F. Ali, Yolov3: An incremental improvement, arXiv abs/1804.02767 (2018) 1804. 02767.
[20] B. Alexa, C.Y. Wang, H.M. Liao, Yolov4: Optimal speed and accuracy of object detection, arXiv:2004 (2020) 10934.
[21] H.S. Bhadauria, M.L. Dewal, Intracranial hemorrhage detection using spatial fuzzy c-mean and region-based active contour on brain CT imaging, Signal Image Video Process 8 (2) (2014) 357–364.
[22] D.M. Alawad, A. Mishra, H. Tamjidul, AIBH: accurate identification of brain hemorrhage using genetic algorithm-based feature selection and stacking, Mach. Learn. Knowl. Extr 2 (2) (2020) 56–77.
[23] M.D. Hssayeni, M.S. Croock, A. Al-Ani, H.F. Al-khafaji, Z.A. Yahya, B. Ghoraani, Intracranial hemorrhage segmentation using a deep convolutional model, Data 5 (1) (2020) 14.
[24] A. Arab, B. Chinda, G. Medvedev, W. Siu, H. Guo, T. Gu, S. Moreno, A fast and fully-automated deep-learning approach for accurate hemorrhage segmentation and volume quantification in non-contrast whole-head CT, Sci. Rep. 10 (1) (2020) 1–12.
[25] J.L. Wang, H. Farooq, K.I. Ali, Segmentation of intracranial hemorrhage using semi-supervised multi-task attention-based U-net, Applied Sciences 10 (9) (2020) 3297.
[26] W.Wu, Y.Lu, R.Mane, C.Guan, Deep learning for neuroimaging segmentation with a novel data augmentation strategy, Annu Int Conf IEEE Eng Med Biol Soc. (2020) 1516-1519.
[27] M. Ganeshkumar, V. Ravi, V. Sowmya, E.A. Gopalakrishnan, K.P. Soman, C. Chakraborty, Identification of intracranial haemorrhage (ICH) using ResNet with data augmentation using CycleGAN and ICH segmentation using SegAN, Multimed, Tools Appl (2022).
[28] Y. Qiu, C. S. Chang, J.L. Wang, L.Ko. Semantic segmentation of intracranial hemorrhages in head CT scans. 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS) (2019) 112-115.
[29] J. Cho, S.Y. Park, M. Karki, E. Lee, S. Ko, J.K. Kim, D. Lee, J. Choe, J. Sun, M. Kim, S. Lee, et al., Improving sensitivity on identification and delineation of intracranial hemorrhage lesion using cascaded deep learning models, J. Digit. Imag 32 (3) (2019) 450–461.
[30] H. Kuang, K.M. Bijoy, Q. Wu, Segmenting hemorrhagic and ischemic infarct simultaneously from follow-up non-contrast CT images in patients with acute ischemic stroke, IEEE Access 7 (2019) 39842–39851.
[31] S. Thay, P. Aimmanee, B. Uyyanavara, P. Rukskul, Fast hemorrhage detection in brain CT scan slices using projection profile based decision tree, In Proc, Int. Con. on Intelligent Information Technology (2018) 18-21.
[32] Q. Zhou, et al., Transfer learning of the ResNet-18 and DenseNet-121 model used to diagnose intracranial hemorrhage in CT scanning, Curr. Pharm. Des 28 (4) (2022) 287–295.
[33] G. Vrbancic, M. Zorman, and V. Podgorelec, Transfer learning tuning utilizing grey wolf optimizer for identification of brain hemorrhage from head CT images, In Proc. 6th Student Comput. Sci. Res. Conf (2019) 61–66.
[34] Y.T. Chen, Y.L. Chen, Y.Y. Chen, Y.T. Huang, H.F. Wong, J.L. Wang, J.J. Wang, Deep learning–based brain computed tomography image classification with hyperparameter optimization through transfer learning for stroke, Diagnostics 12 (4) (2022) 807.
[35] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N.G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, P. Warier, Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, The Lancet 392 (10162) (2018) 2388–2396.
[36] A. Sage, B. Pawel, Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest, Applied Sciences 10 (21) (2020) 7577.
[37] H. Ye, F. Gao, Y. Yin, D. Guo, P. Zhao, Y. Lu, X. Wang, J. Bai, K. Cao, Q. Song, H. Zhang, et al., Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network, Eur Radiol 29 (11) (2019) 6191–6201.
[38] M.A. Ferlin, G. Michał, K. Arkadiusz, M. Agnieszka, S. Edyta, G. Małgorzata, A. Sabis, A comprehensive analysis of deep neural-based cerebral microbleeds detection system, Electronics 10 (18) (2021) 2208.
[39] T.H.Y. Le, A.C. Phan, H.P. Cao, T.C. Phan, Automatic identification of intracranial hemorrhage on CT/MRI image using meta-architectures improved from region-based CNN, World Congress on Global Optimization. Cham: Springer; (2019) 740–750.
[40] T. Lia, Y. Zouc, P. Baiab, S. Lia, H. Wang, X. Chena, Z. Mengc, Z. Kangc, G. Zhouabd, Detecting cerebral microbleeds via deep learning with features enhancement by reusing ground truth, Comput. Methods Programs Biomed 204 (2021), 106051.
[41] A.A. Mohammed, W.R. Kim, E.Y. Kim, Y. Noh, D.H. Kim, Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach, Neuroimage Clin 28 (2020), 102464.
[42] M.J. Myung, K.M. Lee, H.G. Kim, J. Oh, J.Y. Lee, I. Shin, E.J. Kim, J.S. Lee, Novel Approaches to Detection of Cerebral Microbleeds: Single Deep Learning Model to Achieve a Balanced Performance, J. Stroke Cerebrovasc. Dis. 30 (9) (2021), 105886.
[43] O.F. Ertugrul, M.F. Akil, Detecting hemorrhage types and bounding box of hemorrhage by deep learning, Biomed Signal Process Control 71 (2022).
[44] A. Mohiyuddin, A. Basharat, U. Ghani, V. Peter, S. Abbas, O.B. Naeem, M. Rizwan, Breast tumor detection and classification in mammogram images using modified YOLOv5 Network, Comput Math Methods Med 2022 (1–17) (2022) 1–16.
[45] D. Mason, Pydicom: An open source DICOM library, Med Phys 38.6(10) (2011).
[46] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, C. Xu, Ghostnet: More features from cheap operations. IEEE/CVF Conf. Comput. Vis, Pattern Recognit (2020) 1580–1589.
[47] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, IEEE Trans. Pattern Anal. Machine Intelligence 37 (9) (2015) 1904–1916.
[48] S. Liu, L. Qi, H. Qin, J. Shi, J. Jia, Path aggregation network for instance segmentation. IEEE Conf. Comput. Vis, Pattern Recognit (2018).
[49] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection. IEEE Conf. Comput. Vis, Pattern Recognit (2017).
[50] E.P. Reis, F. Nascimento, M. Aranha, F.M. Saco, B. Machado, M. Felix, A. Stein, E. Amaro, Brain Hemorrhage EXtended (BHX): Bounding box extrapolation from thick to thin slice CT images. (version 1.1), PhysioNet (2020).
[51] Tzutalin. Tzutalin/Labelimg. (2018) online: https://github.com/tzutalin/labelImg.
[52] T.Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, Zitnick C L, Dollár P., Microsoft coco: Common objects in context. European Conference on Computer Vision, Springer, Cham, (2014).