Archives of Academic Emergency Medicine. 2025; 13(1): e23

ORIGINAL RESEARCH

Detection of Body Packs in Abdominal CT scans Through Artificial Intelligence; Developing a Machine Learning-based Model

Sayed Masoud Hosseini[1], Seyed Ali Mohtarami[2], Shahin Shadnia[1], Mitra Rahimi[1], Peyman Erfan Talab Evini[1], Babak Mostafazadeh[1]*, Azadeh Memarian[3], Elmira Heidarli[1]

1. Toxicological Research Center, Excellence Center of Clinical Toxicology, Department of Clinical Toxicology, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2. Department of Computer Engineering and Information Technology, (PNU), Tehran, Iran

3. Emergency Medicine, School of Medicine, Mazandaran University of Medical Sciences, Sari, Iran

Received: October 2024; Accepted: November 2024; Published online: 26 December 2024

Abstract: **Introduction:** Identifying the people who try to hide illegal substances in the body for smuggling is of considerable importance in forensic medicine and poisoning. This study aimed to develop a new diagnostic method using artificial intelligence to detect body packs in real-time Abdominal computed tomography (CT) scans. **Methods:** In this cross-sectional study, abdominal CT scan images were employed to create a machine learning-based model for detecting body packs. A single-step object detection called RetinaNet using a modified neck (Proposed Model) was performed to achieve the best results. Also, an angled Bbox (oriented bounding box) in the training dataset played an important role in improving the results. **Results:** A total of 888 abdominal CT scan images were studied. Our proposed Body Packs Detection (BPD) model achieved a mean average precision (mAP) value of 86.6% when the intersection over union (IoU) was 0.5, and a mAP value of 45.6% at different IoU thresholds (from 0.5 to 0.95 in steps of 0.05). It also obtained a Recall value of 58.5%, which was the best result among the standard object detection methods such as the standard RetinaNet. **Conclusion:** This study employed a deep learning network to identify body packs in abdominal CT scans, highlighting the importance of incorporating object shape and variability when leveraging artificial intelligence in healthcare to aid medical practitioners. Nonetheless, the development of a tailored dataset for object detection, like body packs, requires careful curation by subject matter specialists to ensure successful training.

Keywords: Artificial intelligence; Body packing; Tomography, X-ray computed; Diagnostic imaging; Poisoning

Cite this article as: Hosseini SM, Mohtarami SA, Shadnia S, et al. Detection of Body Packs in Abdominal CT scans Through Artificial Intelligence; Developing a Machine Learning-based Model. Arch Acad Emerg Med. 2025; 13(1): e23. https://doi.org/10.22037/aaemj.v13i1.2479.

1. Introduction

The act of concealing illicit drugs within an individual's body is a common method used for transporting small amounts of substances. A commonly observed method entails the ingestion of drug-filled packets, referred to as "body packing", with individuals who engage in this practice being termed "body packers". Substances like opium, cocaine, heroin, amphetamines, ecstasy, and cannabis derivatives like hashish are substances commonly transported by individuals known as body packers. In cases where drug transportation within the body is suspected, imaging procedures are recommended as the most effective means to verify this assumption. While certain indicators, such as the "double-condom" sign, may suggest the presence of drug packages, they are not commonly observed on standard abdominal X-rays. Nowadays, abdominal CT scan is recognized as the most efficient technique for detecting drug packets located in the abdominopelvic region (1-3).

Timely detection of individuals who conceal drugs internally and accurate identification of the location of the drug packets is essential in situations where drug container leakage or rupture is suspected (4). Radiographic interpretation can be complicated by a variety of factors, such as the expertise level of the radiologist, limited contrast resolution inherent to imaging modality, administration of enemas, small size of ingested packets, increased bowel contents and gas, intestinal motility, large urinary bladder stones, intra-abdominal calcifications, and fecal impaction (5). Hence, due to the potential inaccuracies in identifying body packs using radiological imaging and the resulting social and medico-legal implications, an accurate method of identification is essential in this particular scenario.

Medical imaging is a commonly utilized method in digital health for the timely identification and assessment of illnesses. Various techniques, such as Magnetic Resonance

*Corresponding Author: Babak Mostafazadeh; Toxicological Research Center, Excellence Center of Clinical Toxicology, Department of Clinical Toxicology, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Tel/Fax: +98-21-55424041, Email: mstzbmd@sbmu.ac.ir; mstzbmd@yahoo.com, ORCID: https://orcid.org/0000-0003-4872-9610.

Imaging (MRI), X-ray, computed tomography (CT) scan, and positron emission tomography (PET) scans, are employed in this process (6). Recent progress in computer-assisted interventions has demonstrated encouraging outcomes in the realm of medical image analysis (7, 8). The rapid advancement of artificial intelligence (AI) has the potential to enhance medical diagnostics significantly, leading to a transformation in the field by enhancing predictive accuracy, expediting processes, and increasing overall efficiency (9). The utilization of vast datasets, advanced algorithms, and substantial computational capabilities has enabled deep neural networks to be particularly efficacious in tasks related to image analysis and interpretation (10). AI algorithms can examine various types of medical images such as CT scans, MRIs, X-rays, wireless capsule endoscopy (WCE), and ultrasounds, aiding healthcare professionals in more precise, quick identification and diagnosis of various conditions (9, 11).

The convolutional neural network (CNN) stands out as a well-established algorithm within the realm of deep learning models. This class of artificial neural networks has emerged as a prominent technique in computer vision applications following remarkable outcomes in image processing and object recognition competitions, as features can potentially manifest at any location within the image. There has been a notable increase in enthusiasm among researchers in radiology regarding the promising prospects offered by CNN (12). No research has yet explored the application of real-time artificial intelligence for detecting body packs. To address this research gap, this study seeks to employ a single-step RetinaNet object detection approach (utilizing a transformed neck) for identifying body packs in abdominal CT scans, marking the first instance of such investigation.

2. Methods

2.1. Study design and setting

In this cross-sectional study, images were obtained by capturing frames from the abdominal CT scan recordings of individuals with body packs, who were diagnosed and managed by the toxicology and radiology specialists at Loghman Hakim Hospital from March 2019 to February 2023.

Abdominal CT scan images featuring one or more body packs were included in the study.

The study received approval from the ethics committee of Shahid Beheshti University of Medical Sciences, identified by reference number IR.SBMU.RETECH.REC.1401.260. All methods were performed in accordance with the relevant guidelines and regulations by the ethics committee of Shahid Beheshti University of Medical Sciences. General informed consent was obtained from all patients admitted to Loghman Hakim Hospital to use their data anonymously for educational and research purposes. In cases where participants were unable to provide consent themselves, consent was obtained from their immediate family members. The informed consent obtained at our institutions also included authorization for potential future analyses.

2.2. Participants

The dataset encompasses all eligible abdominal CT scan images recorded within the designated study period, thus eliminating the need for a sample size calculation. A total of 888 abdominal CT scan images featuring one or more body packs were included in the study.

The study exclusively incorporated eligible cases for analysis. Subsequently, meticulous detection of all frames extracted from the films was conducted through the creation of angled bounding boxes. The researchers thoroughly scrutinized the images to verify the precision of detection by outlining bounding boxes around the complete air packer.

Inclusion criteria encompassed instances where the presence of a body pack was verified through collaboration between a toxicologist and a radiologist utilizing abdominal CT scans. Additionally, cases involving individuals over the age of 18 who exhibited severe symptoms of substance overdose post-hospitalization, with confirmation of body packing through radiological abdominal CT scans, were considered. Excluded from the study were cases involving body staffers, where drug packages were introduced into the body via anal or vaginal routes.

Following the compilation of a pertinent body packer dataset, it was subsequently segregated into two distinct sets: train and test. The train set underwent training utilizing various object detection models, among which the model proposed in this study was included. This model denoted as Body Packer Detection (BPD) within this research, was a focal point of the investigation.

2.3. Procedure for capturing images

A high-resolution single-lens reflex (SLR) camera, either professional or semi-professional, was utilized for capturing images. The camera had a minimum resolution of 1024×768 pixels. The lens used for capturing images had a macro power of either 60mm or 105mm. All medical personnel involved in the process were thoroughly trained in operating this equipment and doing the following:

P1. Adequately illuminate the surroundings using natural or white light.

P2. Configure the camera to its highest resolution, with a minimum setting of 1024 × 768 pixels.

P3. Position the camera at a distance ranging from 30 to 35 centimeters from the computer monitor.

P4. Ensure that the playback speed of the abdominal CT scan film is set to 30 frames per second (FPS = 30).

P5. Align the vertical orientation of the camera lens perpendicular (90 degrees) to the CT-scan images displayed on the monitor. Hold the camera with both hands in a vertical or horizontal position to maintain alignment.

P6. Confirm that no shadows are present in the area being photographed, ensuring the film is clear, free from shadows, in focus, and at an appropriate distance.

P7. Transfer the recorded videos from the camera to computer files without compromising the file size or image resolution during the transfer process.

Following the capture procedure, the film frames need to be converted into images. Additionally, in compliance with the minimum imaging requirements, videos captured by smartphones, meeting the specified camera conditions, by medical professionals and toxicologists were deemed acceptable for inclusion in this study.

2.4. Datasets

In this investigation, images were obtained by extracting frames from videos captured by researchers using iOS and Android smartphones, adhering to specified conditions outlined in the protocol. The videos processed for analysis were standardized to a frame rate of 30 frames per second (FPS=30). The extraction of images was facilitated through the utilization of VLC software, with a recording ratio set at two-second intervals for automatic screenshot capture. Subsequently, the images were stored in jpg format and manually reviewed to ensure quality and the presence of body pack by two researchers, namely BMZ and SAM.

Images lacking body pack were excluded from the dataset, resulting in a total of 888 images featuring one or more body packs being chosen for further examination. The average dimensions of the images were measured at 850×310 pixels. Software tools were employed to delineate suitable bounding boxes (Bboxes) around the body packs. Bounding boxes are typically defined by two points, commonly representing the top-left and bottom-right corners of the box. These rectangular labels are commonly utilized in tasks related to object detection and localization, offering a clear method to specify the position and dimensions of objects within an image. The accuracy of the Bbox annotations for body packs was verified by three researchers, namely BMS, SAM, and SS.

2.5. Data preparation

Given the diverse sizes and sources of the images, significant pre-processing was required for each image dataset, consuming 80% of the time allocated for preparing a suitable dataset for deep learning applications. Image data comes in various formats, with RGB being a popular choice for color images. The initial step involved generating a collection of images with bounding box (bbox) labels denoting the body pack positions.

Unlike object recognition in natural scenes, detecting body packs presents unique challenges such as scale variations, arbitrary orientations, and dense objects due to their movement within the digestive system. Notably, detecting body packs with arbitrary orientations in abdominal CT scan views posed a specific challenge. The initial concept explored in this study pertains to this issue. Thus, in addition to employing horizontal bounding boxes (Figure 1-a), rotated bounding boxes (Figure 1-b) were utilized for rotated object detection (13, 14). The rotated bounding box approach introduced a fifth parameter, denoted as p angle, alongside the standard four parameters for bbox identification. The value of p angle was determined through the following equation:

$$\theta_P = \begin{cases} \theta & if \quad \theta \leq \pi/2 \\ (\theta - \pi/2) & if \quad \theta > \pi/2 \end{cases}$$

The Morphology Transformation technique was employed to alter the shape and appearance of images with bounding box labels. Initially, the data underwent normalization, a crucial preprocessing step involving the rescaling of pixel values to a specific range. This process is essential for addressing issues such as exploding and the vanishing gradient. In the regression phase of the proposed model, an angle transform was applied to the head region (Head) to introduce angle variations during the normalization process.

The subsequent phase involved augmentation, a technique utilized in data preprocessing for image-based deep-learning applications to enhance the quantity and diversity of the training data. Augmentation was specifically applied to the training set (Figure 2), resulting in a notable enhancement in the mean Average Precision (mAP) of the model when tested on images by toxicology experts for real-time body packer detection from abdominal CT scan images under authentic conditions.

In the examination of body packer images conducted by researchers, a notable observation was the utilization of radiopaque body packs positioned at varying angles to correspond with the movement through distinct segments of the gastrointestinal tract, such as the stomach, small intestine, and large intestine. Consequently, DropBlock, a regularization method for convolutional networks, was employed in the course of this investigation.

2.6. Validation

The assessment of the object recognition model involves identifying all instances of body packs within the images. It is important to note that an image may contain multiple body packs, and the detection process should specifically target body packs while excluding objects from other categories, such as kidney stones or fecal impact, which may require differential diagnosis. In this research, the mean average precision (mAP) metric was utilized, with the data structured like the COCO dataset, using standard bounding box (bbox) parameters (x1, y1, width, height).

Additionally, an angle parameter was incorporated into the bbox parameters in our study, resulting in the following format: x1, y1, width, , height.

In assessing the presence of an object, the Intersection over Union (IoU) metric is employed as a means of determining similarity. This metric is derived by dividing the area of overlap between two bounding boxes by the total area encompassed by their union. IoU is a crucial component in the computation of Average Precision, with mAP serving as a metric that gauges the average precision across all object categories. It is commonly utilized for the evaluation of object recognition models (Figure 3).

Precision and recall were calculated through the bellow equation in which TP(c) TP(c) represents true positive, FP(c) represents false positive, and FN(c) represents false negative. In the case of TP(c) a proposal was made for class c, and there was an object of class c, and in the case of FP(c) a proposal was made for class c, but there is no object of class c and in case of FN(c) no proposal was made for class c, and there was an object of class c.

Precision = TP(c)/TP(c)+FP(c)
Recall = TP(c)/TP(c)+FN(c)

The standard interpretation of Average Precision (AP) involves determining the area under the precision-recall curve. Precision and recall values typically range from 0 to 1, resulting in AP scores also falling within this range. The mAP is computed by averaging the AP values across all classes and/or various IoU thresholds, which may vary based on the specific detection challenges being addressed (15). The AP and mAP metrics are computed using the following formula:

$$AP = \int_0^1 p(r)dr$$

$$mAP = 1/N \sum_{i=1}^{N} APi$$

2.7. Technical evaluation

In this study, Python programming language along with PyTorch, pandas, numpy, and sci-kit-learn libraries were utilized for the development, training, and validation processes. Additionally, a novel model was employed to enhance the outcomes. The objective of the research was to establish conditions conducive to simulating real-world scenarios based on images captured by medical professionals using mobile devices. The aim was to facilitate the practical application of this research in the development of an assistant tool to enhance the precision and reliability of body packer identification. The proposed model underwent rigorous testing during the image preparation and augmentation stages. The dataset was partitioned into 80% for training and 20% for evaluation, following an 80:20 split ratio. The optimization algorithm employed in this model was Adamw. Optimizer = dict(type='AdamW', lr=base_lr, weight_decay=0.05).
Furthermore, the selection of the number of epochs for training was determined to be between 100 and 500, guided by the analysis of the diminishing loss curve and the stabilization observed in the loss curve. The adjustment of the learning rate within the model was made by the following conditions:

```
# learning rate
param_scheduler = [
dict(
type='LinearLR',
start_factor=1.0e-5,
# use cosine lr from 10 to 20 epoch
type='CosineAnnealingLR',
eta_min=base_lr * 0.05,
begin=max_epochs // 2,
end=max_epochs,
T_max=max_epochs // 2,
by_epoch=True,
convert_to_iter_based=True).
```

In this research, all the models underwent training utilizing MM Detection, a tool for object detection, to mitigate potential issues related to model implementation and to provide a standard training pipeline (16).

2.8. Proposed framework (Body Packer Objection Model)

Object detection is commonly assessed through two primary models: one-stage and two-stage detectors (17). One-stage detectors prioritize rapid inference speeds, while two-stage detectors emphasize high accuracy in localization and recognition. The two-stage detection model involves a dual-step process: initially identifying Regions of Interest (RoI) by generating candidate boxes, followed by classifying these RoIs and refining location predictions. The first step utilizes a Region Proposal Network (RPN) to propose RoI candidates, enabling the model to identify potential object regions within the image or video. This network essentially guides the model on where to focus its attention.
Traditionally, methods such as selective search were employed for this purpose, but these were computationally intensive. The advent of RPN significantly enhanced efficiency and speed, thereby reducing the computational burden, particularly in detection networks like Fast RCNN. The two-stage object detection models are commonly referred to as the R-CNN family, with numerous instances of such models available.
Conversely, one-stage techniques involve a single model that partitions the image into regions, revealing bounding box and label possibilities for each region (18-20).
The Faster R-CNN model represents an enhanced iteration of the Fast R-CNN framework, designed to achieve improved computational efficiency. This advancement is achieved by incorporating a convolutional neural network (CNN) as the feature extractor for suggesting rectangular objects during the proposal phase, as opposed to employing a selective search algorithm. The features of the proposed objects are subsequently shared with the detector model, facilitating tasks such as bounding box regression and classification (19).
In two-stage detectors like R-CNN or Faster R-CNN, the initial stage involves a region proposal network (RPN) that diminishes the number of potential object locations and filters out a majority of background instances. Subsequently, in the second stage, classification is performed for each identified candidate object location. Techniques such as adjusting the balance between foreground and background through their respective proportions or employing strategies like Online

Hard Example Mining (OHEM) to select a limited number of anchors per batch are utilized for effective management of object detection processes.

One widely used technique in machine learning is OHEM, which involves the selection of examples that the model confidently predicts as positive but are actually negative. This strategy, referred to as Hard Example Mining aims to identify and prioritize challenging examples for the model.

In certain object detection datasets, easily identifiable examples are prevalent alongside a limited number of difficult instances. The automated identification and inclusion of these challenging examples can enhance the efficiency and effectiveness of the training process. OHEM is a method that adapts Stochastic Gradient Descent (SGD) by sampling examples in a non-uniform manner based on their current loss values.

SGD is an iterative optimization method that utilizes mini-batches of data to estimate the gradient expectation rather than calculating the full gradient using all the available data. OHEM autonomously identifies challenging instances, thereby enhancing the effectiveness and efficiency of training, while eliminating the need for various heuristics and hyperparameters commonly employed in this process (21).

Initially, the RPN module provides a refined selection of boxes with certain backgrounds eliminated. Subsequently, a balance between backgrounds and objects is achieved in the following stage through the utilization of OHEM, such as maintaining a ratio of 3:1.

In single-stage object detectors, a large number of potential object locations are systematically selected within the image, typically around 100,000 locations, to comprehensively cover spatial positions and scales (including scaler and aspect ratios). Additionally, the training process involves utilizing background samples that are readily distinguishable. The issue of imbalance between background samples and object classes during training poses a significant challenge in the accuracy of single-stage detectors, a problem not encountered in two-stage detectors.

In single-stage networks, the generation of numerous bounding boxes makes it impossible to address this imbalance solely through techniques like OHEM and other heuristic approaches (22).

RetinaNet employs a focal loss function, which is a cross-entropy loss that is dynamically scaled. The scaling factor diminishes to zero as the confidence in the correct class grows. This feature enables the automatic reduction of the impact of straightforward examples during the training process, thereby swiftly directing the model's attention towards more challenging instances. The utilization of this focal loss contributes to the enhancement of accuracy in a one-stage framework (23).

Three distinct state-of-the-art object detection models were employed in the study. These included one-stage models such as RetinaNet (23), PAFPN (24), and DropBlock (25), as well as two-stage models like Faster R-CNN (19) and Cascade R-CNN (26). The proposed model was primarily inspired by the RetinaNet model, known for its utilization of focal loss, which has been attributed to enhancing the performance of RetinaNet (23).

2.9. Focal loss function

The cross-entropy loss is defined as follows:

$$CE_{(p,y)} = \begin{cases} -\log(p) & \text{if } y=1 \\ -\log(1-p) & \text{otherwise} \end{cases}$$

The above equation can be written as follows:

$$P_t = \begin{cases} p & \text{if } y=1 \\ 1-p & \text{otherwise} \end{cases}$$

$$CE_{(p,y)} = CE_{(pt)} = -\log(p_t)$$

The Focal Loss function introduces a scaling factor (denoted as coefficient) to the cross-entropy function, which reduces the emphasis on easily distinguishable samples. Consequently, these straightforward instances contribute less to the overall loss (Figure 4).

$$CE(p_t) = -\alpha_t \log(p_t)$$

$$FL(p_t) = -(1-p_t)^{\gamma} \log(p_t)$$

The RetinaNet architecture comprises three fundamental components, namely the Backbone, Neck (FPN), and Head as illustrated in Figure 5-A.

The primary aim of the Feature Pyramid Networks (FPN) is to amalgamate features across diverse scales via a technique referred to as Cross Scale Feature Fusion. In the RetinaNet framework, this feature integration process is executed in a direct manner. An inquiry arises regarding the optimality of this configuration for feature combination across different problem domains. Furthermore, it is pertinent to investigate whether all features hold equal significance in the formation of the output, or if there exists an imbalance in their contributions. Notably, a range of structures have been put forth for the FPN in various research endeavors, as depicted in Figure 5-B (27).

In our proposed framework, ResNet18 was utilized as the backbone architecture. Nevertheless, the majority of ideas aimed at improving network efficiency were put into practice in the intermediate segment, commonly known as the neck. Additional pathways and the integration of Dropblock were incorporated in the neck segment of the network (25), resulting in improved outcomes for body packer detection and accelerated convergence.

Additional paths for disseminating information within neural networks play a crucial role. Modifications were implemented in the neck region to establish information pathways connecting the lower layers with the topmost features, and to create a path between the convolutional layers and the upper layers in both the descending and ascending directions of the hierarchical structure. To enhance the feature set, images were utilized to transfer the output to the head regions

(24).

Given the notable diversity in size observed in the body pack images, the establishment of sub-paths proved to be highly beneficial in enhancing the precision and adaptability of the model in identifying the range of body pack sizes.

Furthermore, by introducing additional pathways in the neck region and taking into account the utilization of the Dropout technique within neural networks, particularly in the Foley-connected layer, more favorable outcomes can be achieved. Dropout involves the exclusion of units (both hidden and visible) within a neural network. This process temporarily eliminates all connections, both forward and backward, associated with the dropped node, thereby generating a modified network architecture derived from the original network. Dropout serves as a straightforward method to mitigate overfitting in neural networks. The fundamental concept involves randomly discarding units, along with their connections, from the neural network during the training phase. This strategy prevents units from excessively co-adapting. Throughout the training, dropout selects from a vast number of distinct thinned networks. By reducing the squared norm of the weights, dropout aids in diminishing overfitting (28).

However, the application of the Dropout technique has limited effectiveness on convolution layers due to the spatial correlation of activation units within these layers, allowing information to still propagate through convolution networks despite Dropout. Consequently, DropBlock, a variant of Dropout that eliminates contiguous regions from feature maps, is employed in convolution networks. This concept was integrated into the proposed neck model for bodypacker detection, resulting in improved outcomes. DropBlock operates as a structured Dropout method where units within a connected region of a feature map are simultaneously deactivated. Implementing DropBlock in skip connections, alongside convolution layers, enhances accuracy. Furthermore, incrementally increasing the number of deactivated units during training enhances accuracy and boosts resilience to variations in hyperparameter selections (25).

Both ideas, namely the integration of DropBlock and modifications in the neck of RetinaNet, were implemented in conjunction with a streamlined backbone architecture, specifically ResNet18, to decrease computational complexity (Figure 6). This strategy was primarily directed towards achieving real-time performance and developing a compact and mobile-friendly model suitable for application design. The intended user base for this model includes toxicologists and emergency physicians in treatment facilities.

3. Results

In this study, the researchers were able to explore a variety of deep architectures by leveraging the user-friendly nature of the Backbone family, exemplified by ResNet in the RetinaNet network. By experimenting with ResNet101, ResNet50, ResNet34, and ResNet18, it was determined that the most favorable outcomes were achieved with the Backbone ResNet18.

To assess the efficacy of the proposed body packer detection model, the mAP was employed as a performance metric. The average mAP value across different Intersections over Union (IoU) thresholds ranging from 0.5 to 0.95 in our model was documented as 45.6, which outperformed alternative single-stage and two-stage models. Furthermore, our model demonstrated superior performance in detecting body pack instances of varying sizes, particularly excelling in identifying cases with small and large dimensions at an IoU threshold of 0.75 in comparison to analogous models, as illustrated in Table 1.

In the model we have put forth, the Average Recall sensitivity at maxDets = 100 outperformed that of alternative models across all scenarios, as indicated in Table 2. It obtained a Recall value of 58.5%, which was the best result among the standard object detection methods such as the standard RetinaNet. The outcomes presented in Tables 1 and 2 demonstrate that our proposed model achieves the highest AUC value relative to other models. Two additional gif files were included in the supplementary information to enhance the visualization of the proposed body packer detection model's performance.

4. Discussion

A thorough examination conducted in 2023, utilizing databases including Web of Science and Scopus focused on the exploration of "body packer detection" through computational methodologies like artificial intelligence and image processing. The review revealed a lack of prior studies employing such techniques, prompting researchers to curate and categorize a dataset specifically for this investigation.

The Bounding Box (Bbox) is a tool commonly employed in computer vision for object detection. However, while this method is effective in various scenarios, it may not be the most suitable for detecting certain objects, such as body packers, which are angled rectangles oriented in different ways. In our research, we utilized the display of angled rectangles (oriented bounding boxes) to precisely locate body packs in the training dataset, drawing parallels to the work of Etten et al. (15) who employed spherical Bbox for identifying spherical objects like red blood cells in medical applications. Similarly, Jiang's study (29) utilized a dataset containing angular objects to enhance image analysis, resulting in outcomes akin to our research findings.

This research employed RetinaNet with focal loss to analyze the effectiveness of abdominal CT scans in detecting body packers. Several other studies have also been conducted in this field, including research by Paul F. Jaeger et al. on the integration of RetinaNet and U-Net networks, known as Retina U-Net. This study demonstrated improved performance in diagnosing malignant or benign lesions in lung CT scan medical images, achieving an mAP10 score of 35.8%, surpassing results obtained by other networks (30). Ke Yan et al. also investigated the use of a cannulation network to detect im-

paired regions in CT scans, employing a three-dimensional network based on Convolutional Neural Networks (CNN) as opposed to traditional two-dimensional approaches, resulting in enhanced outcomes compared to detectors such as Faster-RCNN (31). San-Gil Lee and colleagues utilized a modified Single Shot MultiBox Detector (SSD) to predict hepatic lesions in CT scans, achieving an average precision score of 53.3% in identifying such lesions (32). Additionally, Zihao Li and colleagues introduced the concept of utilizing a multi-view feature pyramid network (FPN) for interpreting hepatic lesions, which demonstrated a 2.91% higher mean average precision (mAP) compared to the baseline approach. In their model, a 3D MVP-Net derived from the primary CT-scan data was integrated into the system, and through innovative techniques, the features were amalgamated to provide a comprehensive analysis of the lesion using R-CNN and RPN algorithms (33). In another study, Ethan H. and co-authors proposed a novel circular representation for medical object detection, presenting CircleNet as an anchor-free detection framework in lieu of traditional bounding boxes. Their circle-based approach exhibited enhanced detection performance and greater rotation invariance when identifying glomeruli and nuclei in pathological images (34).

Abdominal CT scans were found to exhibit higher sensitivity and specificity than abdominal X-rays due to their enhanced contrast clarity.

Abdominal computed tomography (CT) imaging without contrast reveals the presence of body packs as numerous oval or circular foreign objects dispersed throughout the abdomen, emitting radiation (35, 36). These objects exhibit hyperdensity on abdominal CT scans, typically falling within the 20-70 Hounsfield units (HU) range. Radiologically, body packs manifest in various forms, including the "egg bag" or "tic-tac" sign denoting the presence of multiple uniformly dense objects with distinct borders, the "double condom" sign indicating air trapped between layers, the "rosette" sign representing air trapped within package knots, the "halo" sign characterized by a bright rim encircling the package, the "black crescent" sign showing a crescent of air surrounding the package, the "air sign" depicting a transparent triangle between closed objects, and the "feces-like" appearance resembling sharp-layered feces between the body packaging and intestinal wall (37, 38). The diverse shapes and sizes of these objects pose a diagnostic challenge, prompting the initiation of this study as a preliminary step towards leveraging artificial intelligence for object detection to aid in diagnosis.

Various body pack forms and sizes have prompted researchers to seek a more adaptable machine-learning framework. To address this, the network architecture was modified to allow for flexible adjustments in the network hierarchy, establishing connections between the Feature Pyramid Network (FPN) outputs as proposed by Wang et al. (13). The transmission of information within neural networks is crucial for object recognition, with low-level features being particularly important for detecting large objects. However, the lengthy path between high-level and low-level features in the feature pyramid poses challenges in accurately localizing large objects. To address the issue of recognizing body packers of varying sizes, one approach involves enhancing the feature pyramid with precise low-level positional data and reducing the information path. An enhancement to the bottom-up path based on FPN, known as PAFPN, has been introduced to refine the feature pyramid architecture and shorten the information path. PAFPN facilitates information flow between lower and higher network layers by establishing sub-paths and enhancing network accuracy (24).

To evaluate the model's ability to detect body packers, a specific regularization technique known as DropBlock was employed, resulting in improved accuracy of the model on the body packers dataset (25, 39). The authors utilized this approach to enhance the model's performance by generating diverse shapes and perspectives of body packs as they traverse various regions of the digestive system during abdominal CT scans.

The suggested framework demonstrates adaptability, offering significant potential to enhance model accuracy through the utilization of more intricate CNNs as the foundation and diverse RPNs. In striving to achieve real-time capabilities for detecting body packers in abdominal CT scans, the researchers deliberately refrained from escalating computational demands.

5. Limitations

The main limitation of this study is the small sample size for image processing using deep learning algorithms. Our research was conducted exclusively with data from a single center, Loghman Hakim Hospital. To enhance the applicability of results, future studies should consider enlarging the sample size or incorporating data from various hospital or provinces.

6. Conclusions

This research utilized a deep learning network to detect body packs in abdominal CT scans, demonstrating the significance of considering object shape and diversity while exploiting artificial intelligence in the medical field to support healthcare professionals. However, the creation of a specialized dataset for object detection, such as body packs, necessitates meticulous curation by domain experts for effective training.

Moreover, the development of extensive detection datasets holds promise in maximizing the capabilities of deep learning models for object detection within both artificial intelligence and clinical contexts. The precise identification facilitated by this approach enhances the practical utility of intelligent assistants in real-world clinical settings. Future endeavors will focus on expanding the dataset to encompass thousands of samples and incorporating differential diagnoses. Enhancements in data fine-tuning and the adoption of advanced object detection techniques are anticipated to en-

hance the overall outcomes of the study.

7. Abbreviations

AI: artificial intelligence
Bbox: oriented Bounding box
BPD: Body Packer Detection
CAD: computer-aided diagnosis
CNN: convolutional neural network
CT: computed tomography
FN(c): false negative
FPN: Feature Pyramid Network
FPS: frames per second
HU: Hounsfield units
IoU: intersection over union
mAP: mean average precision
MRI: Magnetic Resonance Imaging
OHEM: Online Hard Example Mining
PET: positron emission tomography
RoI: Regions of Interest
RPN: Region Proposal Network
SGD: Stochastic Gradient Descent
SLR: single-lens reflex
SSD: Single Shot MultiBox Detector
FP(c): false positive
TP(c): true positive

8. Declarations

8.1. Acknowledgments

The present work was supported by the Toxicological Research Center, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences.

8.2. Funding

This research was supported by the research project, funded by the Toxicological Research Center, Shahid Beheshti University of Medical Sciences.

8.3. Conflict of interest

The authors declare no competing interests.

8.4. Authors' contributions

Azadeh Memarian and Mitra Rahimi: Data gathering and Resources. Sayed Masoud Hosseini: Data curation and Writing – review & editing.
Seyed Ali Mohtarami: Choose Models/Algorithm, Training & Evaluate Models and Visualization. Babak Mostafazadeh and Peyman Erfan Talab Evini: Methodology. Shahin Shadnia: Conceptualization and Writing – review & editing and. Elmira Heidarli: Writing – original draft. All authors read and approved the final version of manuscript.

8.5. Data availability

Data will be made available on request.

8.6. Using artificial intelligence chatbots

None.

References

1. Puntonet J, Gorgiard C, Soussy N, Soyer P, Dion E. Body packing, body stuffing and body pushing: Characteristics and pitfalls on low-dose CT. Clinical Imaging. 2021;79:244-50.
2. Pinto A, Reginelli A, Pinto F, Sica G, Scaglione M, Berger FH, et al. Radiological and practical aspects of body packing. The British journal of radiology. 2014;87(1036):20130500.
3. HASANIAN MH, ABOU ALMASOUMI Z. Consequence of body packing of illicit drugs. 2007.
4. Pinto A, Reginelli A, Pinto F, Sica G, Scaglione M, Berger FH, et al. Radiological and practical aspects of body packing. British Journal of Radiology. 2014;87(1036).
5. Reginelli A, Russo A, Urraro F, Maresca D, Martiniello C, D'Andrea A, et al. Imaging of body packing: errors and medico-legal issues. Abdominal Imaging. 2015;40(7):2127-42.
6. Ganatra N, editor A Comprehensive Study of Applying Object Detection Methods for Medical Image Analysis. 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom); 2021 17-19 March 2021.
7. Pinto-Coelho L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering (Basel). 2023;10(12).
8. Li M, Jiang Y, Zhang Y, Zhu H. Medical image analysis using deep learning algorithms. Front Public Health. 2023;11:1273253.
9. Al-Antari MA. Artificial Intelligence for Medical Diagnostics-Existing and Future AI Technology! Diagnostics (Basel). 2023;13(4).
10. Ya-ting F, Qiong L, Tong X. New Opportunities and Challenges for Forensic Medicine in the Era of Artificial Intelligence Technology# br. Journal of Forensic Medicine. 2020;36(1):77.
11. Nakada A, Niikura R, Otani K, Kurose Y, Hayashi Y, Kitamura K, et al. Improved Object Detection Artificial Intelligence Using the Revised RetinaNet Model for the Automatic Detection of Ulcerations, Vascular Lesions, and Tumors in Wireless Capsule Endoscopy. Biomedicines. 2023;11(3).
12. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights into Imaging. 2018;9(4):611-29.
13. Wang Y, Zhang Y, Zhang Y, Zhao L, Sun X, Guo Z. SARD: Towards scale-aware rotated object detection in aerial imagery. IEEE Access. 2019;7:173855-65.
14. Tang T, Zhou S, Deng Z, Lei L, Zou H. Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks. Remote Sensing. 2017;9(11):1170.

15. Van Etten A, editor Satellite imagery multiscale rapid detection with windowed networks. 2019 IEEE winter conference on applications of computer vision (WACV); 2019: IEEE.
16. Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X, et al. MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:190607155. 2019.
17. Sultana F, Sufian A, Dutta P. A review of object detection models based on convolutional neural network. Intelligent computing: image processing based applications. 2020:1-16.
18. Girshick R, Donahue J, Darrell T, Malik J, editors. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.
19. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems. 2015;28.
20. Carranza-García M, Torres-Mateo J, Lara-Benítez P, García-Gutiérrez J. On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. Remote Sensing. 2020;13(1):89.
21. Shrivastava A, Gupta A, Girshick R, editors. Training region-based object detectors with online hard example mining. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
22. Tian Z, Chu X, Wang X, Wei X, Shen C. Fully convolutional one-stage 3d object detection on lidar range images. Advances in neural information processing systems. 2022;35:34899-911.
23. Lin T-Y, Goyal P, Girshick R, He K, Dollár P, editors. Focal loss for dense object detection. Proceedings of the IEEE international conference on computer vision; 2017.
24. Liu S, Qi L, Qin H, Shi J, Jia J, editors. Path aggregation network for instance segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
25. Ghiasi G, Lin T-Y, Le QV. Dropblock: A regularization method for convolutional networks. Advances in neural information processing systems. 2018;31.
26. Cai Z, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence. 2019;43(5):1483-98.
27. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S, editors. Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
28. Nitish S. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15:1.
29. Jiang Y, Zhu X, Wang X, Yang S, Li W, Wang H, et al. R2CNN: Rotational region CNN for orientation robust scene text detection. arXiv preprint arXiv:170609579. 2017.
30. Jaeger PF, Kohl SA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer H-P, et al., editors. Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection. Machine Learning for Health Workshop; 2020: PMLR.
31. Yan K, Bagheri M, Summers RM, editors. 3D context enhanced region-based convolutional neural network for end-to-end lesion detection. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I; 2018: Springer.
32. Lee S-g, Bae JS, Kim H, Kim JH, Yoon S, editors. Liver lesion detection from weakly-labeled multi-phase ct volumes with a grouped single shot multibox detector. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11; 2018: Springer.
33. Li Z, Zhang S, Zhang J, Huang K, Wang Y, Yu Y, editors. MVP-Net: multi-view FPN with position-aware attention for deep universal lesion detection. Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22; 2019: Springer.
34. Nguyen EH, Yang H, Deng R, Lu Y, Zhu Z, Roland JT, et al. Circle representation for medical object detection. IEEE transactions on medical imaging. 2021;41(3):746-54.
35. Taheri MS, Moharamzad Y, Nahvi V. Abdominal CT findings of ruptured opium packets in a body packer. European Journal of Radiology Extra. 2009;70(1):e21-e3.
36. Shahnazi M, Taheri MS, Pourghorban R. Body packing and its radiologic manifestations: a review article. Iranian Journal of Radiology. 2011;8(4):205.
37. Tsang HKP, Wong CKK, Wong OF, Chan WLW, Ma HM, Lit CHA. Radiological features of body packers: An experience from a regional accident and emergency department in close proximity to the Hong Kong International Airport. Hong Kong Journal of Emergency Medicine. 2018;25(4):202-10.
38. Niewiarowski S, Gogbashian A, Afaq A, Kantor R, Win Z. Abdominal X-ray signs of intra-intestinal drug smuggling. Journal of Forensic and Legal Medicine. 2010;17(4):198-202.
39. Yelleni SH, Kumari D, Srijith P. Monte Carlo DropBlock for modeling uncertainty in object detection. Pattern Recognition. 2024;146:110003.

Table 1: Average Precision (AP) and mean AP for body packer detection

Methods	Backbone	AP0.50:0.95	AP50	AP75	APs	APM	APL
Two-stage							
Faster-rcnn _fpn	ResNet50	41.7	85.8	35.3	41.4	43.2	46.7
Cascade-rcnn_ sac_Detectron	ResNet50	42.4	85.7	36.4	42.8	42.2	36.8
One-stage							
RetinaNet _fpn	ResNet50	42.3	85.7	35.6	43.1	40.2	40.4
RetinaNet _fpn	ResNet18	42.3	86.7	36.7	42.8	41.5	46.7
RetinaNet _ FPN_DropBlock	ResNet18	43.7	85.5	37.6	43.3	44.9	56.7
RetinaNet _PAFPN	ResNet18	42.7	85.5	37.0	43.4	40.8	46.8
Proposed model*	ResNet18	44.6	85.7	41.5	44.7	43.9	58.9
Proposed model*+oBbox	ResNet18	45.6	86.6	42.4	45.8	45.9	59.8

Intersection over union (IoU) threshold =X: APX (Example: IoU threshold =50%: AP50); APS: Small objects are defined as being between $0\hat{2}$ and $32\hat{2}$ pixels in area; APM: Medium objects are defined as being between $32\hat{2}$ and $96\hat{2}$ pixels in area; APL: Large objects are defined as being between $96\hat{2}$ and $1e5\hat{2}$ pixels in area; oBbox: oriented bounding boxes.

Table 2: Average recall (AR) for Body package detection

Methods	Backbone	AR	ARs	ARM	ARL
Two-stage					
Faster-rcnn _fpn	ResNet50	50.1	59.7	51.6	46.7
Cascade-rcnn_ sac_Detectron	ResNet50	51.6	51.9	51.0	36.7
One-stage					
RetinaNet _fpn	ResNet50	53.2	53.0	54.1	40.0
RetinaNet _fpn	ResNet18	53.1	52.8	53.9	46.7
RetinaNet _ FPN_DropBlock	ResNet18	53.3	52.8	54.8	56.7
RetinaNet _PAFPN	ResNet18	52.6	52.6	52.5	46.7
Proposed model* (our model1)	ResNet18	57.7	57.0	59.9	60.0
Proposed model*+oBbox (our model2)	ResNet18	58.5	58.1	60.1	62.1

Intersection over union (IoU) threshold =X: APX (Example: IoU threshold =50%: AP50); APS: Small objects are defined as being between $0\hat{2}$ and $32\hat{2}$ pixels in area; APM: Medium objects are defined as being between $32\hat{2}$ and $96\hat{2}$ pixels in area; APL: Large objects are defined as being between $96\hat{2}$ and $1e5\hat{2}$ pixels in area; oBbox: oriented bounding boxes.

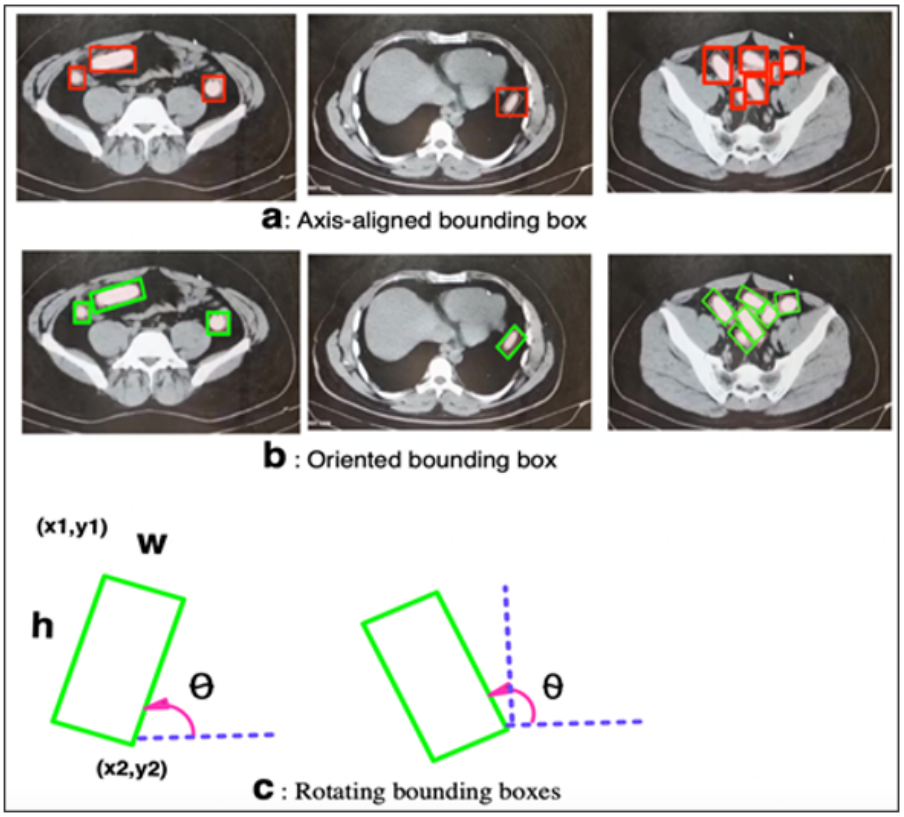

Figure 1: Object detection model with oriented bounding boxes.

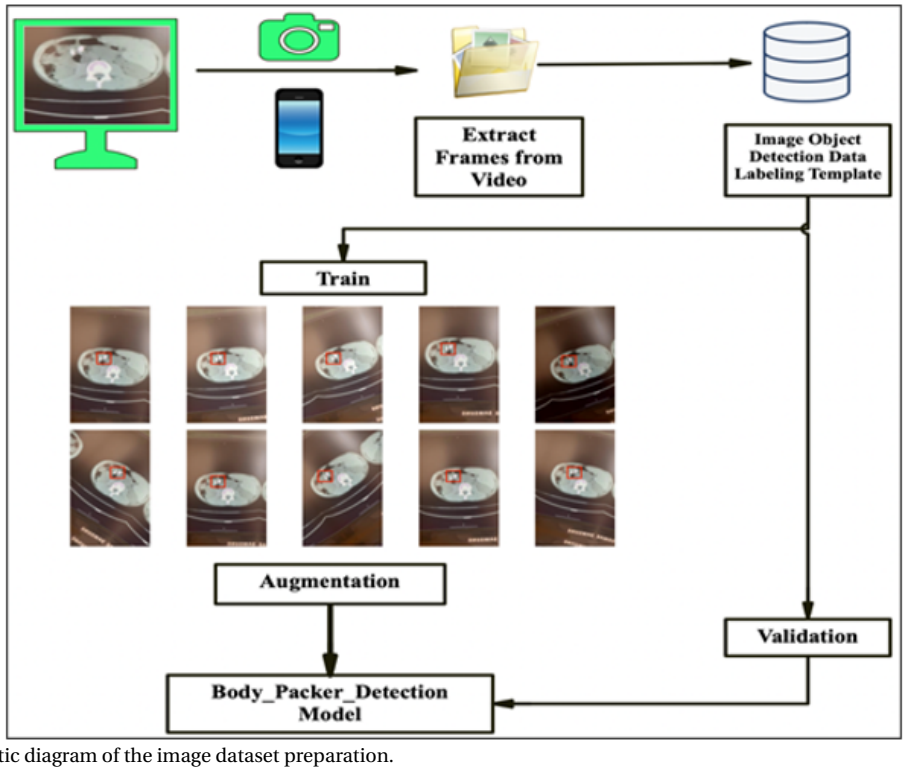

Figure 2: Schematic diagram of the image dataset preparation.

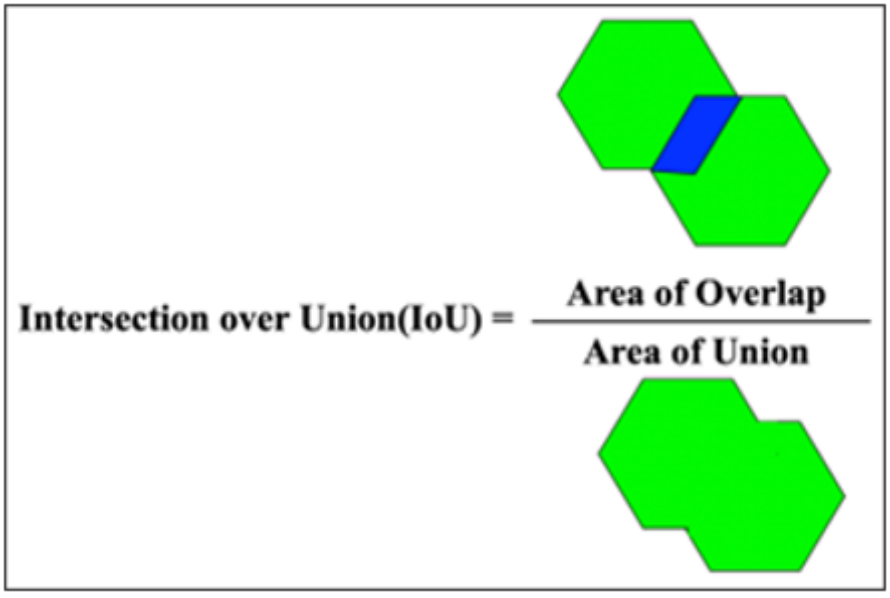

Figure 3: Intersection over Union (IoU).

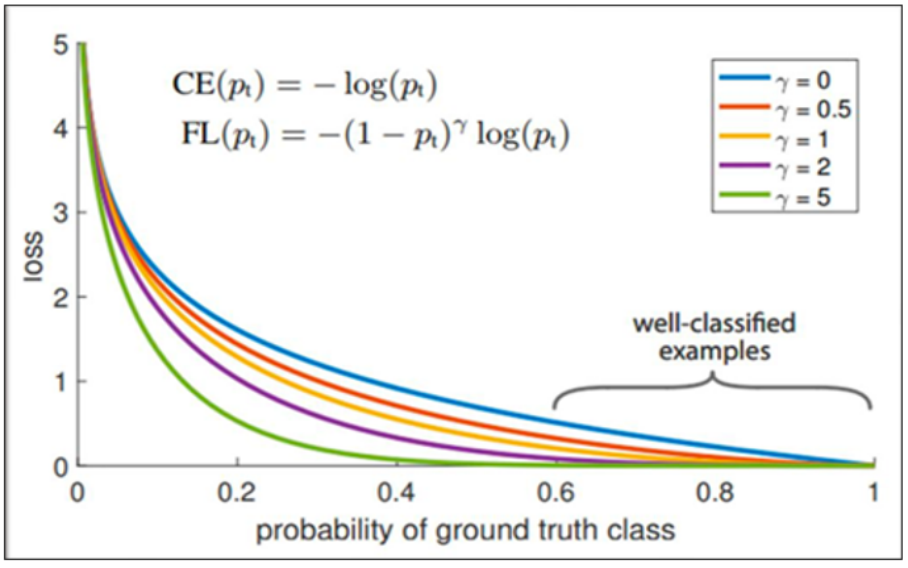

Figure 4: The Focal Loss introduces a modification to the conventional cross entropy criterion by incorporating a factor of (1 pt). When is set to a value greater than zero, it diminishes the loss proportion for accurately classified instances (pt > .5), thereby emphasizing more on challenging, misclassified instances. The parameter denotes the focusing parameter that directs attention towards difficult misclassified instances, while represents the balancing coefficient as suggested in the primary literature (23).

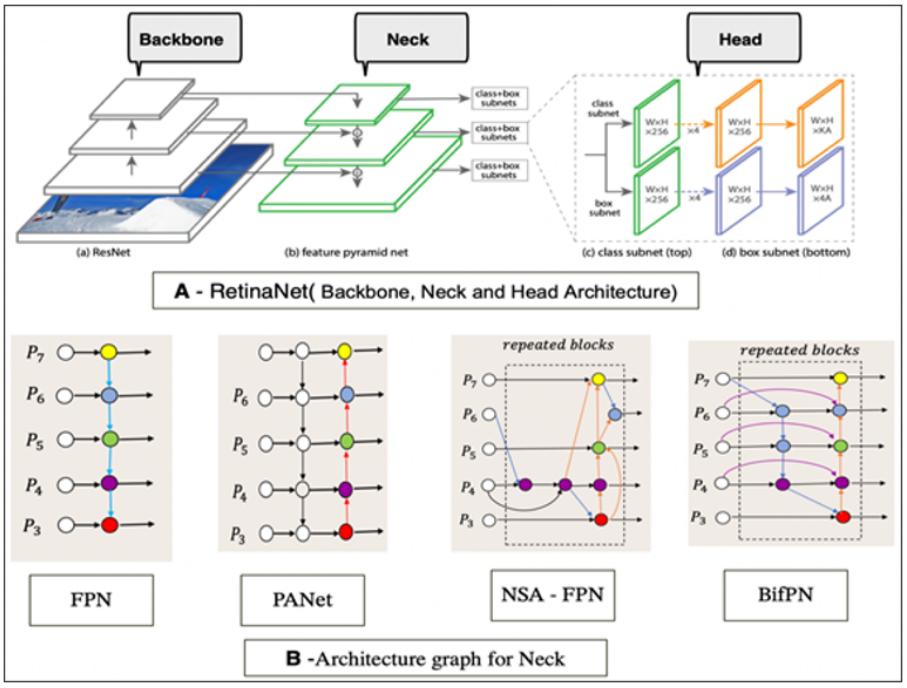

Figure 5: (A) The RetinaNet architecture utilizes the ResNet backbone network in conjunction with the Feature Pyramid Network (FPN) as the feature extractor, while two additional Convolutional Neural Networks (CNNs) are responsible for classification and regression tasks. (B) Various neck architectures can be explored within the RetinaNet framework.

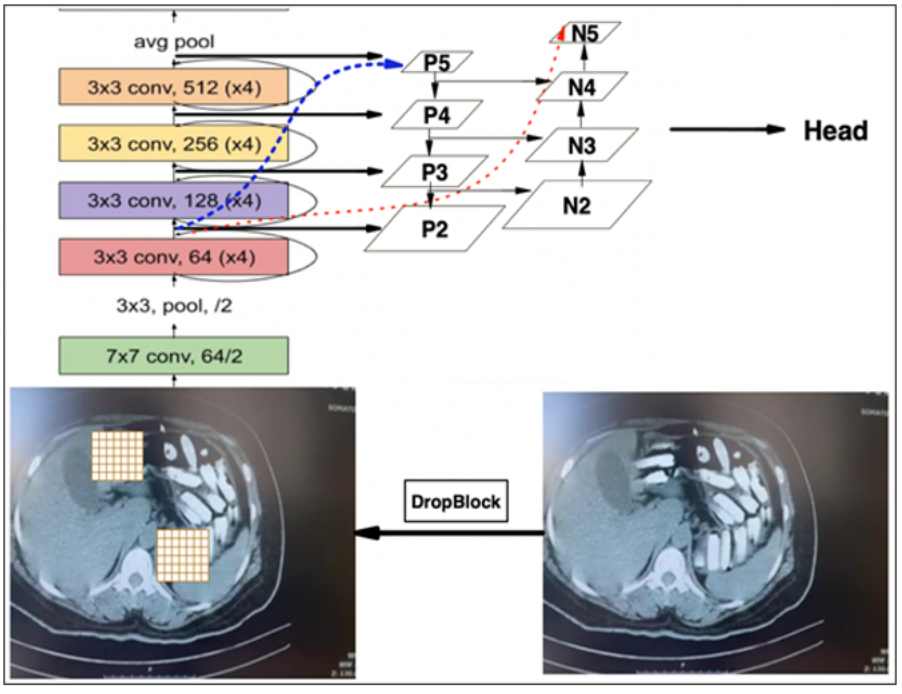

Figure 6: In the proposed model, both concepts, namely the integration of DropBlock and modifications in the neck of RetinaNet, were implemented in conjunction with a streamlined backbone architecture, specifically ResNet18, to decrease computational complexity.