International Journal of Advanced Technology and Engineering Exploration, Vol 9(96)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2021.875911

Research Article

Context based healthcare informatics system to detect gallstones using deep learning methods

Veena A[1*] and Gowrishankar S[2]
Research Scholar, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru –560056, India(Affiliated to VTU, Belagavi – 590018)[1]
Professor, Department of Computer Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru – 560056, India(Affiliated to VTU, Belagavi – 590018)[2]

Received: 03-May-2022; Revised: 15-November-2022; Accepted: 18-November-2022

Abstract
Gallstone is a chronic condition that affects people around the globe. Gallstone disease (GSD) is a major challenge on healthcare systems across the globe, and it is the most widespread diseases among people with abdominal discomfort who are admitted to the emergency rooms. The size of the gallbladder organ ultrasound is affected by a multitude of features, including the layer of the gallbladder. As a result, gallbladder scans from various parameters make it a difficult process. In this paper, we have developed a healthcare informatics system to investigate and detect the gallstones present. A detailed comparison of different algorithms inspired by cutting edge models for object detection is used in this paper. These include faster region-based convolutional neural network (Faster-R-CNN, mask regional convolutional neural network (Mask R-CNN), and single shot detector (SSD). The suggested model, which is based on the mask R-CNN, SSD, Faster R-CNN technique, distinguishes stones in the gallbladder by extracting suggestions from the stone area. The Mask R-CNN model was modelled and trained using various backbone networks. We have used the ultrasound images for the experiment obtained from the medical professionals. The ultrasound images gathered are focused on various contexts such as gender, age, and people from urban and rural areas. The results indicate that the Mask R-CNN with backbone network of Resnet-101-FPN combination outperforms in case of object detection.

Keywords
Deep learning, Streamlit, SSD, Faster R-CNN, Mask R-CNN, Healthcare, Informatics system.

1.Introduction
Gallbladder stone ailment is the utmost prevalent gastrointestinal illness that necessitates hospitalization, with a projected 800,000 cholecystectomies in the United States each year [1].

The Gallbladder is a digestive organ that stores and concentrates bile, a digestive material that is produced into the small intestine to facilitate digestion. In the gallbladder, gallstones are deposits of hardened digestive fluid. Cholecystitis is when the gallbladder becomes inflamed. Gallbladder difficulties are caused by a blockage in the bile ducts, which are tubes that carry bile between the gallbladder, liver, and small intestine.

*Author for correspondence

Cholecystitis is often due to a blockage caused by gallstones. Other factors that can cause Cholecystitis includes tumors, infections, or issues with blood circulation.

Gallstones affect over 25 million residents in the United States, with women accounting for 65% to 75% of those get affected. Gallstones are “silent” in most cases, meaning they don’t cause any noticeable signs. Gallstones or other gallbladder problems may cause sharp agony in the upper right or middle of the abdomen, the agony that extends or radiates to the right shoulder or back, a tender abdomen that is sensitive when touched, fever, shivering, nausea and vomiting, and jaundice. Gallstone disease occurrence is strongly linked to age and gender, with metabolic risk features for gallstone disease differing between women and men. Gender is thought to be a contextual factor in gallstone disease [2]. Women are at a greater risk as compared to men [3−5]. As

progesterone slows the emptying of the gallbladder, estrogen causes cholesterol to build up in the bile and gallstones to develop. Gallstones affect the female hormones. Obesity is also a contributing factor, since fatter bodies contain more estrogen hormones. Rapid weight loss, on the other hand, raises the danger of low-calorie diets interfering with bile production and, as a result, causing more cholesterol crystallization and gallstone development. Gallstones are more prone to form in patients who have diabetes or any other illness that causes gallbladder contractions or intestinal motility to be reduced, such as an anatomical structural injury. Finally, some research suggests that some people are susceptible to gallstone development due to heredity factors [3−5]. Population-based trial research in men show gallbladder stone illness is expounded to an increased danger of organic impotence. [6]. The most significant risk factor for gallbladder cancer is gallstones. Gallstones differ widely in size. Some individuals have a single huge gallstone, while others form hundreds of tiny gallstones. Gallstones range in size from 5 to 10 mm in diameter as shown in *Figure 1*.

Figure 1 Image showing multiple gallstones

If the gallstone is not treated early, it can lead to serious complications. It may lead to a gallbladder tear or a bile infection. Hence, this project's motivation is to collaborate with physicians, and we do not intend to replace them, but to assist them with the diagnosis.

For a very long time, the healthcare industry has been a pioneer in technical developments and has reaped important benefits. By focusing on the needs of the patients, the digitization of the healthcare sector has assisted the healthcare service providers in developing a stable and critical system. Digitization has aided technology such as artificial intelligence (AI), which can assist medical practitioners in analyzing massive amounts of data in order to gain information and to make well informed decisions. Healthcare is transitioning to a cloud-based model. Thanks to AI and data science, access to a person's medical history would be simple and is not restricted to a specific place or timeframe.

AI has helped telehealth allowing the users to chat live with doctors through online portals. Natural language processing, speech recognition, machine learning, AI-optimized hardware, virtual agents, and deep learning (DL), to name a few areas, have all seen significant advancements. DL is used in an extensive variety of areas [7] including healthcare. DL can take the healthcare sector to the next level by streamlining the procedures and increasing the customer loyalty. Automatic image processing using machine learning and DL approaches has indeed demonstrated potential for segmentation, tissue reconstruction, regression, classification, segmentation, and regression using ultrasonography [8, 9]. Cost reductions would be realized because of the increased efficiency. DL has alternative methods for automatically extracting image characteristics without making any conclusions about the underlying mechanism.

The existing technology used in recognition of these gallstones is very costly and we plan to democratize the technology behind these gallstones by making it as an open-source software. To the best of my knowledge, this article looks at gallstones from a variety of DL algorithmic perspectives, and no other algorithm has been used to compare this gallbladder dataset.

The remainder of the article is structured as discussed below. Section two converses existing surveys and their criticism. Datasets and transfer learning methods used in the diagnosis of gallstones is addressed along with the results in section three and four. Section five discusses a systematic analysis of various methods. Observations and an outlook on the gallstones are discussed in this section. Finally, it is concluded in section six.

2.Literature review

In recent years, several DL algorithms have made major advances in computer vision and healthcare domain. The extension of mask regional convolutional neural network (Mask R-CNN) has been implemented/proposed by Anantharaman et al. [10]. It is an innovative convolutional neural network (CNN) technique for target recognition and

localization, to the domain of dental anatomy. It has been observed that the increasing accuracy on a tiny dataset is a difficult undertaking because the dataset is sparse.

An effective liver tumor segmentation approach based on Mask R-CNN with Resnet 101 is proposed by Haq et al. [11]. The procedure is focused on the Mask R-CNN technique to successfully recognize the tumors in the liver by producing proposals about the tumor area. The differences in the liver and complex tumor regions' shapes, sizes, and locations, make it difficult for Mask region-based convolutional neural network (R-CNN) for segmenting the tumor region.

The Hsieh et al. [12] propose the use of Mask R-CNN to eliminate the noise from the context and find malignant micro calcification (MC's) from clusters. An inception V3 algorithm is employed to differentiate between normal and cancerous MC clusters. It has been observed that cluster grouping precision, MC's naming, and benign and malignant analysis is 93%, 91% and 95% respectively. The accuracy, specificity and sensitivity of the models are also very good. The limitation of this article is that breast cancer stages could have been determined.

Single shot detector (SSD) is implemented by Sha et al. [13] for locating and detecting fractures in the spine (lumbar fracture, cervical fracture, thoracic fracture). The test results show that the scheme's mean average precision (mAP) is 78% and the total time of each identification is 0.048 sec. It is observed that DL requires a lot of time to train CT pictures and labelling the CT images is a laborious task.

Region-based convolutional neural network is implemented by Sokolova et al. [14] used to segment pixel and to locate iris. The output of convolutional neural network is compared with various backbone networks on a manually labelled image dataset composed from various eye operations. It is detected that common objects in context (COCO) weights were used to initialize the backbone networks. It is also observed that there is a slight overfitting in the model. A new object detection model that takes occlusions into account in road scenes is suggested by Kim et al. [15]. A multiple object bounding box (OBB) critical framework was suggested to handle occlusions. The bounding box is projected for feature encoding networks to encrypt latent obstruction and artefact characteristics. It is discovered that the proposed method outperformed both the baseline object detection system and the sophisticated methods.

The Chen et al. [16] demonstrated a technique for implying SSD-mobilenets using reduced edge computing internet of things (IoT) unit. The authors have successfully implemented SSD-mobilenets on the Raspberry Pi 3 using the neural compute stick (NCS). The three datasets validate the proposed system (Pascal VOC, KITTI and LPS2017). The experiments revealed that using four NCS's on SSD mobilenets computed images in 1.7 sec and accelerate video sequences to 9.2 frames per second. It is found that the hardware is a hurdle to pretrain a neural network model to operate on inexpensive IoT device. An end-to-end architecture which uses the cutting-edge Mask R-CNN model for identification and localization is suggested by the Chen et al. [17] to segment and localize the geographical borders. It is observed that the proposed system achieves accurate identification and segmentation performance in Xiamen Island. In future the authors have planned to investigate how village borders evolve over time.

Lin et al. [18] have employed a Fast R-CNN as the object delineation recognition method and AdaBoost as an indicator to improve the precision of the results. It is revealed that the images were annotated with an annotator tool and the SURF method is used to extract the image characteristics. Findings show that the target identification correctness of Fast R-CNN is 91.6 percent in the picture collection with a resolution of 866×652 (pixels), and AdaBoost as the remaining detector improves the precision to 96.76 percent but the time cost is too high. Mohedano et al. [19] suggested the Electroencephalogram (EEG) tool for relevance feedback and contrasted it to the conventional "click – based" mechanism for an item extraction assignment. It is discovered that the EEG-based approach achieved comparable accuracy. It is laborious and mentally taxing to manually annotate images with a mouse, especially in a visual retrieval setting.

The surgical tool is segmented and localized using a spatiotemporal deep network by the Kanakatte et al. [20]. The cholec80 dataset has been used to compare the presentation of the projected technique with the most recent image-based instance segmentation technique. With positive outcomes, it is also contrasted with techniques reported in the works that use frame-level occurrence detection and spatial detection.

Jain et al. [21] presented an abdominal multi-organ segmentation using a fully computerized, registration-free u-net architecture. The usefulness of the suggested method is evaluated using 50 abdominal computed tomography data sets, and it outperforms more established two-dimensional (2D) segmentation techniques in terms of outcomes. This method completely relies on only u-net which is the shortcoming and is trained and tested on 50 images.

Obaid et al. [22] developed a DL framework to aid in the diagnosis of biliary atresia using sonographic images of the gallbladder. Visual geometry group (VGG)16, inceptionV3, ResNet152, and mobilenet are four different types of DL models that are used, and it is observed that mobilenet performs best with accuracy, specificity, and sensitivity values of 97.87%, 97.51%, and 98.18%.

This article by the Qifang et al. [23] employed the SSD and the Faster R-CNN models to identify the aircraft in the high-quality remote sensing image. The test findings revealed that the representations can successfully be used to create a high-resolution distant recognizing image. It is found that both the models can reliably detect the aircraft in a solo act, but cannot reliably detect in a dual act. The test findings of the Faster R-CNN outline in the complex scene are considerably sophisticated than the SSD model, and the test results reveal that the Faster R-CNN typical has a key benefit in the identification of the minor airplane.

Gbcnet is proposed by Basu et al. [24] to categorize gallbladder cancer. To handle spurious textures in gbcnet while minimizing texture biases, the authors propose a curriculum modelled after human visual acuity. Experimental findings show that the gbcnet performs much better than the up-to-the-minute convolutional neural network (CNN) models and uses grayscale images.

A simple feature-enhanced convolutional neural network is recommended by the Tao et al. [25] to recognize low-attitude soaring substances with high precision in real time. An improved characteristic processing component is used to increase the model's capability to extract features. The potential for low-altitude object identification is demonstrated by the suggested method's recognition rapidity of 147 frames per second and mAP of 90.97%. It is observed that the model does not consider a complex environment.

Song et al. [26] implemented multi-scale characteristic learning technique to merge the existing characteristics and intangible features. It is found that the model fuses the features in the network to create the feature pyramid. It is also observed that as and when the parameters increase, the network becomes too deep and the computation time also increases. It is found that the model effectively represents the information about small object features when compared to the widely used backbone network.

Liu et al. [27] propose a novel network architecture to detect small objects and creates multilayer feature maps. It is found that there is an increase in the magnitude of the characteristics by lowering the number of layers. In this manner, meaningful characteristics of varying sizes are obtained without the use of additional elements. It is observed that the architecture does not consider speed as the performance metric.

Rizos and Kalogeraki [28] proposed a DL approach to identify items (such as fish) in underwater photos using two CNNs. The approach used the first CNN for picture classification and the second CNN uses transfer learning. In a cohesive underwater wireless sensor network system, the suggested technique has been put into practice and tested. It is observed that the model is trained and tested on a small dataset.

The swine transformer is implemented by Jung et al. [29], to identify the food. It is described that the purpose of the model is to estimate the performance of the UECFOOD dataset in comparison to existing food object detection models. The results demonstrate that the suggested model outperforms earlier research. It is noticed that the model requires pixel level predictions for transformers which is the main disadvantage.

An object identification novel strategy is implemented by the Xu [30]. It is seen that the synthetic aperture radar is the main strategy. The results found that the receiving antenna's position is constantly shifting as the vehicle moves for more accurate object recognition. The results show that constantly shifting the antenna position is an added overhead.

Park et al. [31] describe a practical approach to identify ghost regions, stolen things, and abandoned objects in closed-circuit television footage. It is surveyed that the two primary tactics used in this

methodology are object segmentation and removal of immobile objects. The main concept is to consider both the present and the past by determining if a potential stationary object from the backdrop model. The findings demonstrate that the segmented item comparable to the model calculation can be found in either the current video frame or the background frame before it.

Multitask learning model is implemented by the Reddy and Kandasamy [32] and is trained with skeleton localization and skeleton scale prediction. It is observed that the pixel's status as a skeleton pixel is determined by skeleton localization, and each skeleton pixel's scale is determined by skeleton scale prediction. The results demonstrate that, despite having a large feature vector, the proposed model outperforms the traditional model.

Based on the above literature, none of the approaches have been used to detect the gallstones. We use advanced models to detect gallstones. The healthcare informatics system is created by using the models such as SSD, Faster R-CNN and Mask R-CNN.

3.Methodology

A.Datasets

The ultrasound images of gallstones were used for experimentation. The data is collected from BGS Global Hospitals, Bengaluru, Nikhara Diagnostic Center Bangalore, and Deenanath Mangeshkar Hospital and Research Center, Pune. The images collected were gallstones that are larger than 2cm. Some images contained single gallstone, some with multiple gallstones and they were all larger than 2cm. The location, extent and degree and nature of calcification are the key features considered when the computerized tomography (CT) images were collected. The data was obtained from reputable medical professionals, and the privacy of the patients who were made aware of the images being collected was protected by anonymizing the images. The images collected are based on different context like gender, age and people from urban and rural areas. We collected around 30 images, each from urban and rural areas. The images gathered included both single and multiple gallstones of various sizes. Among the suspects, 40 (66.6%) had numerous stones, while only 20 (33.4%) had just one. This data collection includes normal CT images, muddy CT images, and images of granular stones; no stones with CT images are included. The raw images, thus obtained go through a sequence of pre-processing stages. To enhance the precision of the images, pre-processing such as resizing, grayscale, auto alignment, and contrast adjustments are performed on the raw images. Flip, rotate, brighten / darken, harvest, shear, blur, and apply random noise to images to maximize the training data. The images are then annotated using tools that are specific to the size and shape of the gallstones.

In recent decades, CNN's have revolutionized pattern recognition. As a result, we can also use CNNs in a diversity of tasks, as well as object detection and image recognition.

TensorFlow object detection application programming interface (API) is a software application for performing object detection functions. For predictions, the TensorFlow object detection API and other related API's (Keras RCNN, YOLO) use pre-trained CNNs within the frameworks. These models are trained to detect a certain set of object categories. Using various datasets, we can retrain these models to recognize custom artefacts.

We have deployed models for detecting the gallstones for our dataset. The models used in this experiment are SSDs, Faster R-CNN and Mask R-CNN [33−35].

B. Models

1. Single shot detectors (SSD)

SSDs are one stage object locator and built on a feed-forward CNN as shown in *Figure 2*. Instead of proposals, a predefined set of default object placements, scales, and aspect ratios is established. During the estimation time, each object category receives scores from the network for the presence of objects in each default box and then modifies the box to recover their object form. You only live once (YOLO) [33]: You look only once is an amalgamated real – time object detection system [36]. YOLO: You look only once is an illustration of SSD.

SSD can use any baseline architecture as the feature extractor, for example mobilenet. The features are saved as feature maps. Since feature maps reflect the image's prominent features at different scales, running multi-box on numerous feature maps increases the probability of any object being identified, localized, and correctly categorized. It is a fully convolutional model, and it also adds multiple scale predictions. Instead of creating a single grid, the SSD creates several grids for different scales. It makes use of feature maps with various levels of resolution to predict objects. Training the model

entails selecting a range of default detector boxes and weights, as well as the difficult mining and data enhancement techniques [34, 35]. The multi-box [36−39] goal served as the foundation for the SSD training target, which has now been expanded to include a number of object groups.

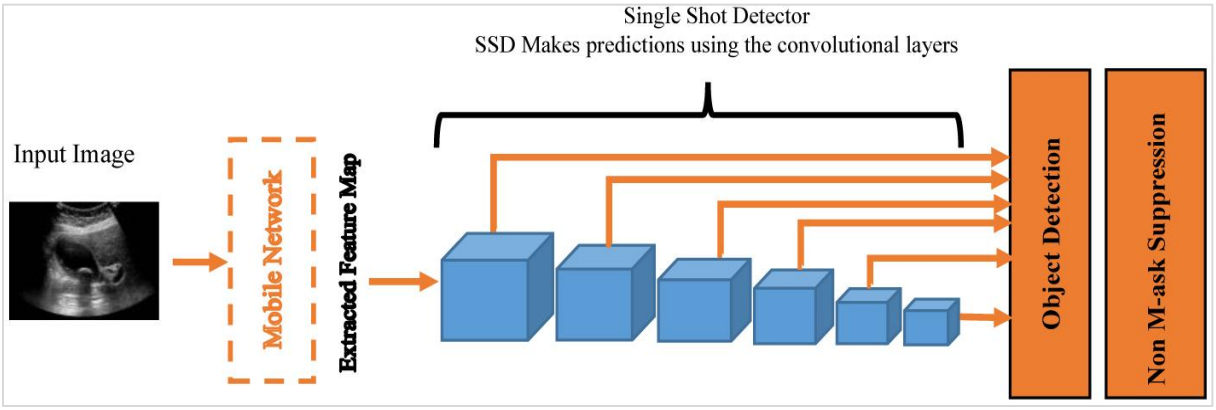


Figure 2 Single shot detector architecture for object detection

Let $x_{ij}^{p} = \{1,0\}$ can be a predictor for matching the i[th] default box to the j[th] boundary box of group p. The total target loss function is the weighted average of the localization and the confidence losses as shown in Equation 1.

$$L(x,c,l,g) = \frac{1}{N}(L_{conf}(x,c) + \alpha L_{loc}(x,l,g)) \quad (1)$$

Where N is the number of matched default boxes, and the localization loss is the smooth L1 loss between the predicted box (l) and the boundary box (g) features. The offsets are regressed for the bounding box's height, width, and center. The weight term α is adjusted to 1 by cross validation (tf), and the confidence loss is the softmax loss across several groups confidences (c).

Smaller artefacts do worse on SSD since they may not be represented in all feature maps. Increasing the input image resolution helps, but it does not solve the problem entirely.

2.Faster region based convolutional neural network (R – CNN)

As shown in *Figure 3*, Faster R-CNN is an object identification technique that progresses on Fast R-CNN [40] by fusing the CNN model with a region proposal network (RPN). Full-image convolutional capabilities are shared by the RPN and the detection network, enabling practically free area recommendations.

RPN is a small network with a convolutional feature map that slides over. This RPN accepts an input and then classifies the appropriate region as object or non-object in a degenerate bounding box location using the special window (n=3RPN). Localization information is provided by the position of the sliding window in reference to the images. Box regression, with relation to this sliding window, gives finer localization information. A collection of k – object proposals is specified for each sliding window location. In terms of size and aspect ratio, each proposal is unique. Anchors are a type of proposition like this. Anchors make it easier to handle objects of various size and aspect ratios. Anchors are essentially various sizes and aspect ratios of sliding windows. When training of RPN, the anchor is noted as positive sample if Intersection of Union (IoU) > 0.7 or IoU is maximum for all anchor box to the ground truth box, and if IoU is less than 0.3 and the anchor is designated as a negative sample, the box regression is trained to degenerate the positive sample box to the ground truth box. There will be 60× 40 sliding window positions and nine anchors for each window position for an image of 1000×600 pixels using VGG by network as feature extractor, resulting in a total of 21500 suggestions. Faster R-CNNs can be developed with a single network and four losses throughout - RPN Regression, RPN Classification, Fast R-CNN classification loss across classes and Fast R-CNN regression loss to predict the boundary between the proposed box and the actual box.

RPN has a tough time dealing with items of varying sizes. Because RPN has fixed receptive fields, tiny things can reside in a very small section of the amenable arena, while huge objects can fill only a portion of the receptive field. The RPN can be trained

for various scales to overcome this problem. Each RPN will use a divergent convolutional layer or group of tiers as input, resulting in a variable-sized receptive field. Small and big objects will be detected with great accuracy. As a result, a single faster region CNN model can concurrently recognize objects ranging in size from tiny to huge.

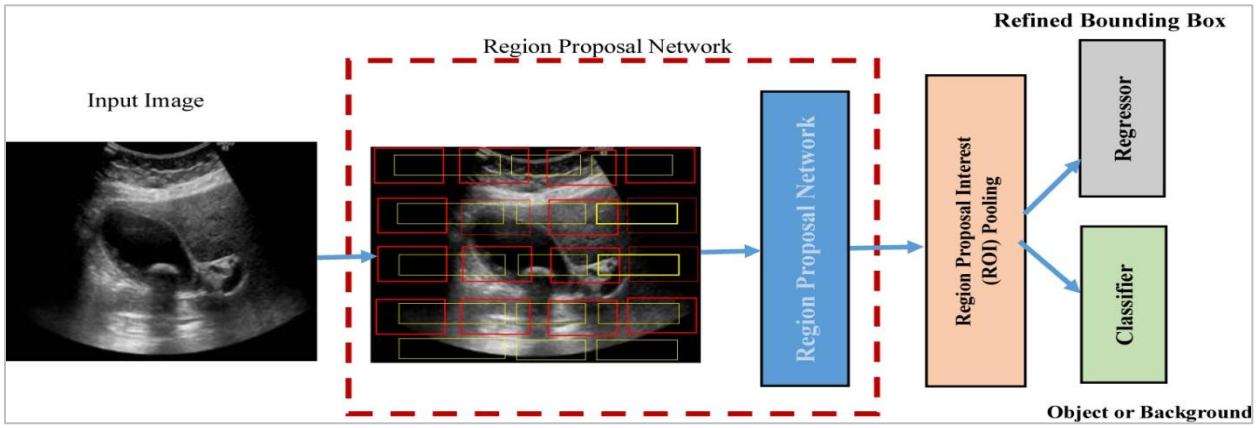


Figure 3 Object Detection using a Faster R-CNN Architecture

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* \; L_{reg} \; (t_i, t_i^*) \quad (2)$$

In Equation 2, p_i denotes the probability that anchor i is an object, the ground truth table for determining if anchor i is an entity is p_i^*.

The ground truth coordinates are designated by t_i*, the estimated four parameterized coordinates are marked by t_i, the normalization terms N_{cls} and N_{box} are set to the number of anchor points and the λ balancing parameter guarantees that both terms L_{cls} and L_{box} are nearly equally valued.

3 Mask region based convolutional neural networks

For example, segmentation and object detection, Mask R-CNN [41] is a widely used DL technique and the schematic illustration of Mask R-CNN architecture is shown in *Figure 4*. It is an instance localization and segmentation method, capable of performing pixel-level segmentation as well as target recognition [38]. Mask R-CNN on region of interest pooling (RoIs) is a hybrid of Faster R-CNN and fully convolutional network (FCN). It uses Faster R-CNN, which in turn uses a convolutional network, ResNet 101 architecture to retrieve the feature maps from the pictures. The candidate bounding boxes are then returned after passing these feature maps via RPN. RPN essentially predicts whether or not an object is present in a given area. And, on these candidate bounding boxes [39], a RoI pooling layer is added to get all of the candidates to the same size. Finally, the candidate bounding boxes are fed into a completely connected layer, which categorizes and guesses the bounding boxes for objects.

Mask RCNN generates the segmentation mask. The computation time is decreased by calculating the area of attention first. With the ground truth boxes, the intersection over union (IoU) is considered for all forecasted regions. IoU is calculated using the formula given below in Equation 3.

$$IoU = \frac{Area\ of\ the\ Intersection}{Area\ of\ the\ Union} \quad (3)$$

If the IoU is larger than or equal to 0.5, the area is measured to be of interest; otherwise, the region is ignored.

Once the RoI is calculated based on IoU, the mask branch is added to the prevailing architecture. The segmentation mask for each area that includes an object is returned by the mask branch. For each region, it returns a mask of size 28×28 that is then increased in size for inference.

The multi-task loss function of Mask R-CNN integrates the segmentation, classification and localization mask losses as shown in Equation 4.

$$L_{max} = L_{cls} + L_{box} + L_{mask} \quad (4)$$

Similar to Faster R-CNN, L_{cls} and L_{box} are used. L_{mask}, which stands for binary cross-entropy loss,

only employs the k-th mask when the area aligns with the ground truth class k.

$$L_{max} = -\frac{1}{m^2} + \sum_{1 \le i,j \le m} \left[y_{ij} \log y_{ij}^{k} + \left(1 - y_{ij}\right) \log\left(1 - y_{ij}^{k}\right) \right] \quad (5)$$

In Equation 5, y_{ij} stands for a cell's (i, j) label in the factual disguise for an area of size m by m, and y_{ij}^{k} stands for the cell's predicted value in the learnt mask for the ground truth class k. There are some of the missing objects, false mask and failure in recognition of objects due to failure in detection filter segmentation and not annotating the images properly.

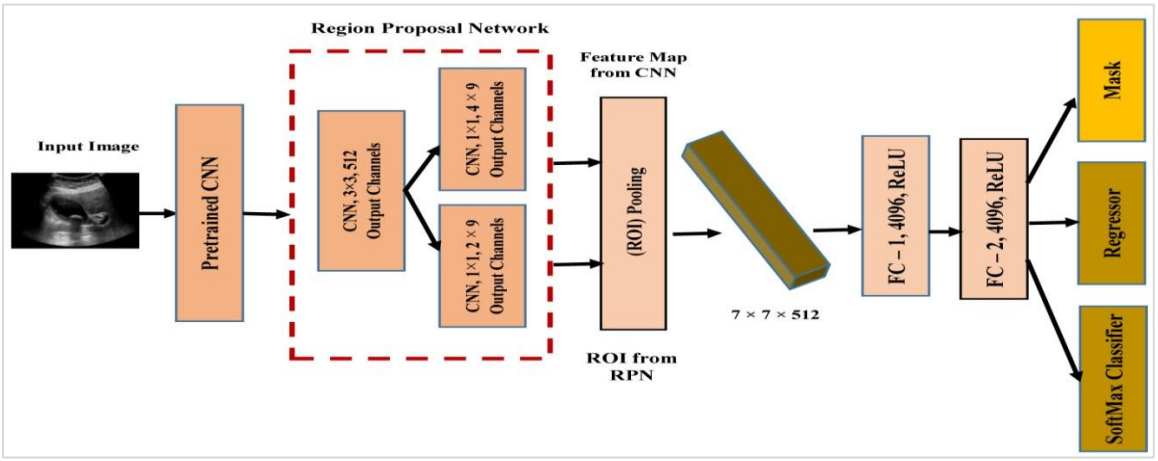


Figure 4 Mask R-CNN architecture for object detection

C. System description

DL's potential is to develop high-level interpretations directly from data without the intervention of domain specialists is one of the key reasons for its popularity. The deep neural network has many non-linear concealed layers. Each layer receives the preceding layer's output and absorbs the data in order from bottom to top [42]. Healthcare informatics system is a system that involves development, analytics, and aggregation of data to meet the needs of the medical industry, as well as the prediction of illness before it happens using the medical data of the patient and his or her families [43]. In this paper, we have developed a health informatics system to investigate the presence of gallstones and the construction of the system is as revealed in the *Figure 5* and *Figure 6* illustrates the sequence diagram of health informatics system.

Healthcare informatics system consists of two modules. In the first module, we have developed the informatics system using streamlit web library and the second part is the parameters are passed to the DL modules to be executed. Streamlit is an open-source framework that converts python scripts into web applications and allows them to deploy instantly. An object detection model is built to detect the gallbladder stones. Then the object detection model is integrated with the streamlit application to make predictions and to visualize those predictions at a given confidence level.

To begin, a python file is used to construct the streamlit web application. This file, contains the code for reading the image and the name of the DL model as input. SSD, Faster R-CNN and, Mask RCNNs are the three DL models used in the streamlit web application. The input images must be annotated or labelled before the model is built. There are many types of annotations available and we have chosen the one that best suits the DL model. Annotations include bounding boxes, polygonal segmentation, semantic segmentation, 3D cuboids, key points and landmarks, and lines and splines, among others. Bounding boxes annotation is employed for SSD and Faster R-CNN models and polygonal segmentation is used for Mask R-CNN model. Objects in the images are often irregular and coarsely shaped, for example medical images. Polygonal segmentations are a form of data annotation in which complex polygons are utilized to precisely specify the shape and placement of the item. LabelImg and VGG annotator tools are used to annotate the input images along with the help of doctors for domain knowledge.

The input data is then separated into two categories: training and testing data in proportions of 80% and 20%, respectively. Jupyter notebook was used to

build and train the Faster R-CNN, Mask R-CNN, and SSD models. The pickle file is then generated from the trained DL models by dumping the model into the pickle file. Pickle is a module that offers binary serialization and de-serialization methods for Python object structures. Unpickling transforms a byte stream (from a binary file or bytes-like item) back into a Python object hierarchy. Pickling transforms a Python object hierarchy to a byte stream [44]. Pickle file is used by the health informatics system to execute the models.

After the health informatics system received the inputs, it then opens the corresponding pickle model file and read the contents of the pickle file in the model variable. Using the model variable, a function is called by passing the arguments such as the image and the model name to the function. The function executes according to the architecture briefed in section 3.2 and sends the results back to the calling function. Healthcare informatics system uses streamlit. Image () function to display the image in the front end along with the bounding boxes / masked region along with their confidence scores.

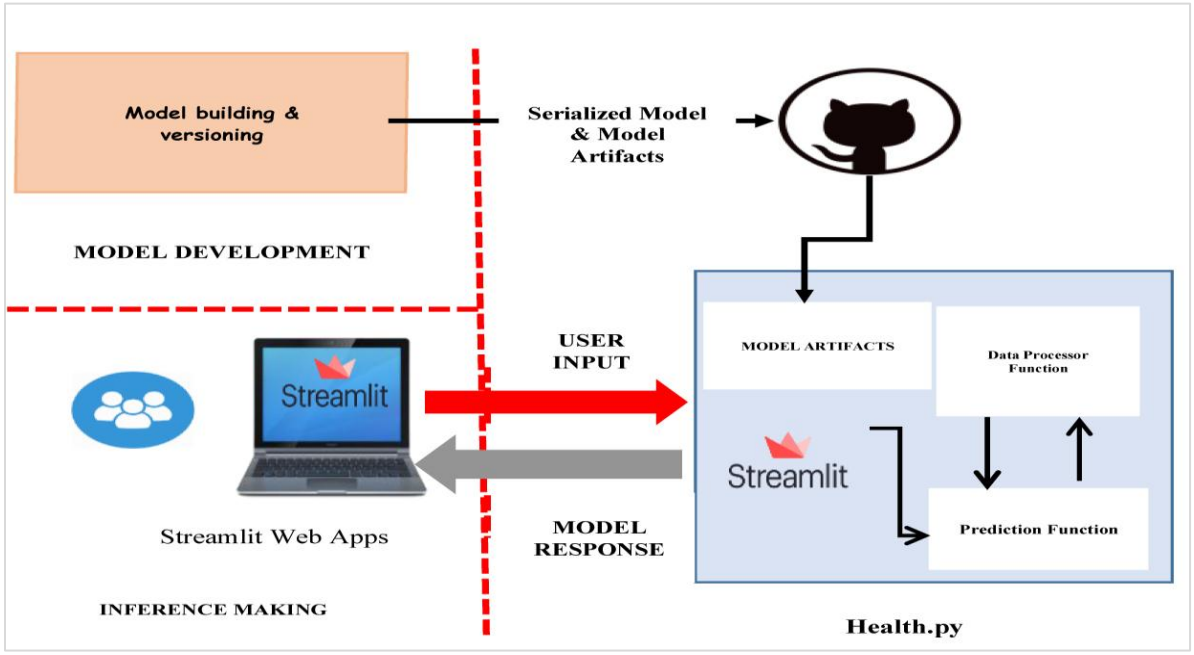


Figure 5 Shows architecture of health informatics system

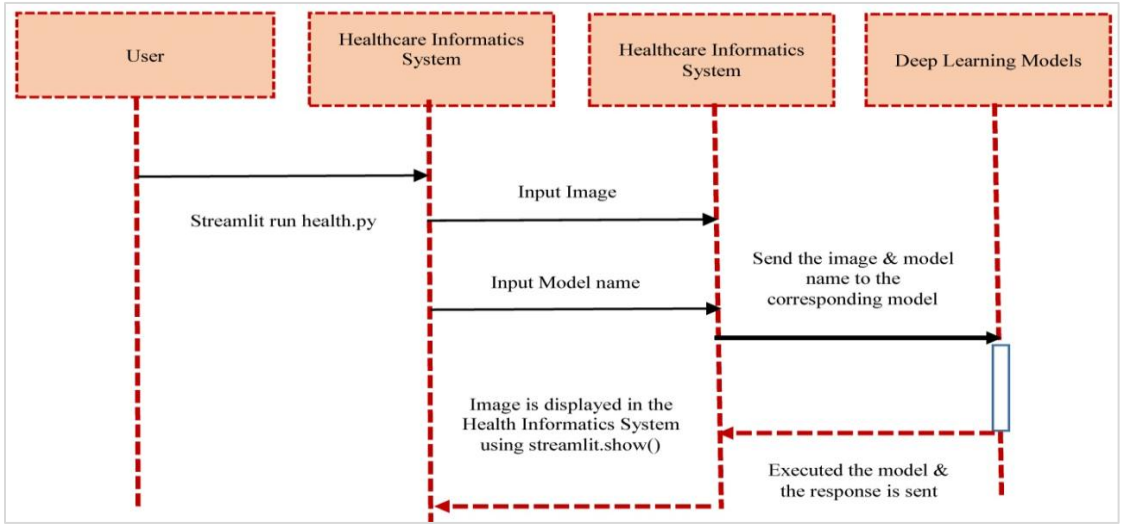


Figure 6 Workflow of the healthcare informatics system

4.Results

The health informatics system is used to investigate gallstones as objects and tabulate the performance using different models. This section displays the model's output. The experimentation is performed on the gallbladder datasets. First, the SSD model was executed to perform gallstone detection. *Figure 7* displays the outcomes of the SSD-efficientDet model in streamlit application with the confidence score of 0.50.

The Google TensorFlow object detection API is non-proprietary programming framework for developing, training, and deploying object detection models that is based on Tensorflow [45]. Using the Tensorflow's object detection API framework, about 100 images were trained on the SSD efficientDet, Faster R-CNNs and, Mask R-CNNs models. Python 3.8.5 and TensorFlow 2.4.1 were used to train the models on the Nvidia GEFORCE GTX 1080 graphics card. With the support of TensorBoard [46], the bounding box and classification loss, we show the values for SSD efficientdet in *Figure 8*, *9* and *10*.

Localization detects the discrepancy between the anticipated boundary box and the actual truth box. The localization loss in *Figure 8* shows that, as the model is trained the loss decreases, increasing the expected predicted boundary box accuracy. On training results, regularization is a tool for reducing model complexity and overfitting. Weights may become too high during model preparation, influencing performance estimation by overpowering the rest of the weights. The model loses its generality and becomes overfit on the training results. Regularization is added to the total loss function.

Uncompleted laparoscopic and open gallbladder removal results in residual gallstones. Gallstone remnants are frequently seen in the common bile duct, cystic duct, and gallbladder. With improved technology or laparoscopic surgery, it is rare for the gallstones to remain. If gallstone residual is found, after complaining of the symptoms, CT scan is done. These CT images are then given to the health informatics system for the detection of the gallstones. Open or laparoscopic surgery is done based on the condition of the patient.

Figure 11 displays the detection of the gallstones in streamlit application using the Faster R-CNN model with a confidence score of 0.90. *Figure 12* shows the output in streamlit application using the Mask R - CNN model with a confidence threshold of 0.75. Some of the configuration parameters used in Mask R-CNN are as follows.

1. Backbone network of ResNet101
2. Batch size of 1
3. Steps per epoch 10
4. Classification FC layers size is 1024
5. Loss weights and Batch Size of 1

A complete list of abbreviations is shown in *Appendix I.*

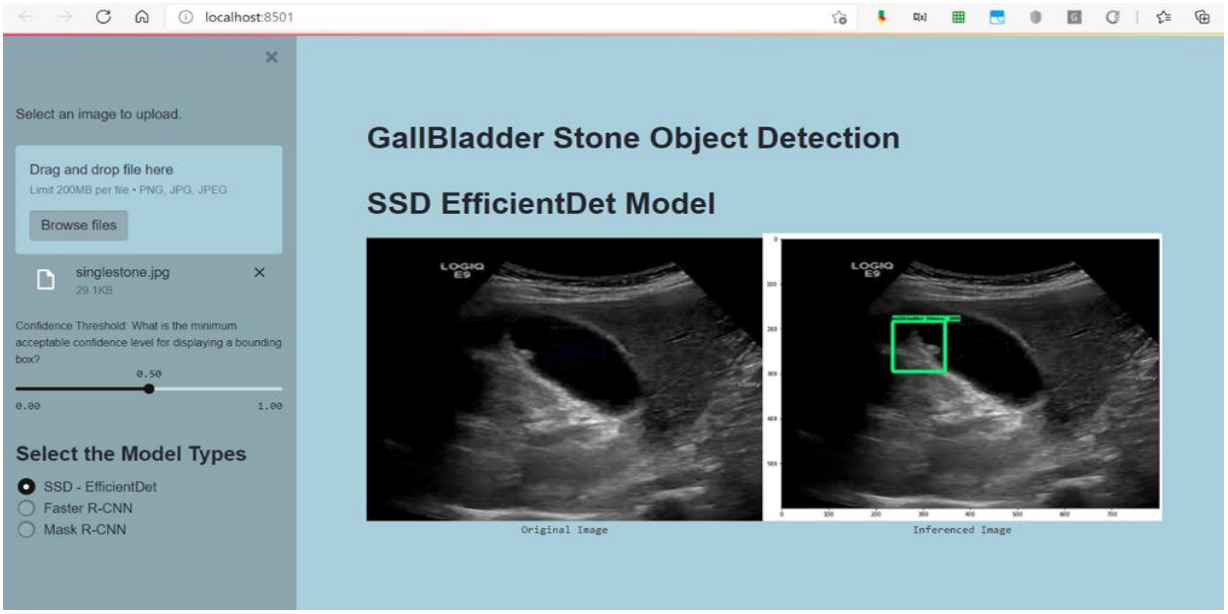


Figure 7 Results of efficientDet from the healthcare informatics system

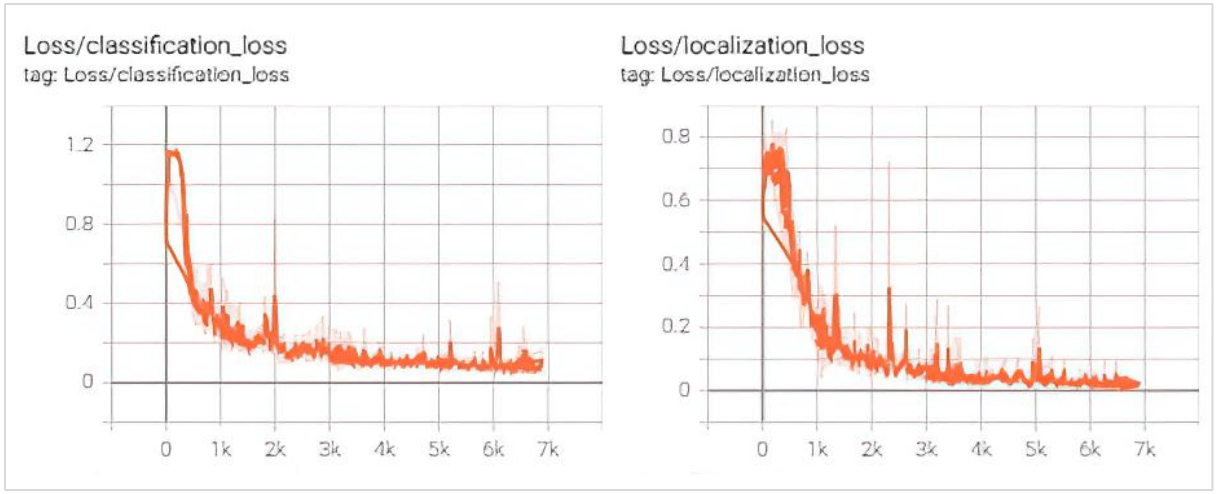


Figure 8 Classification and localization losses

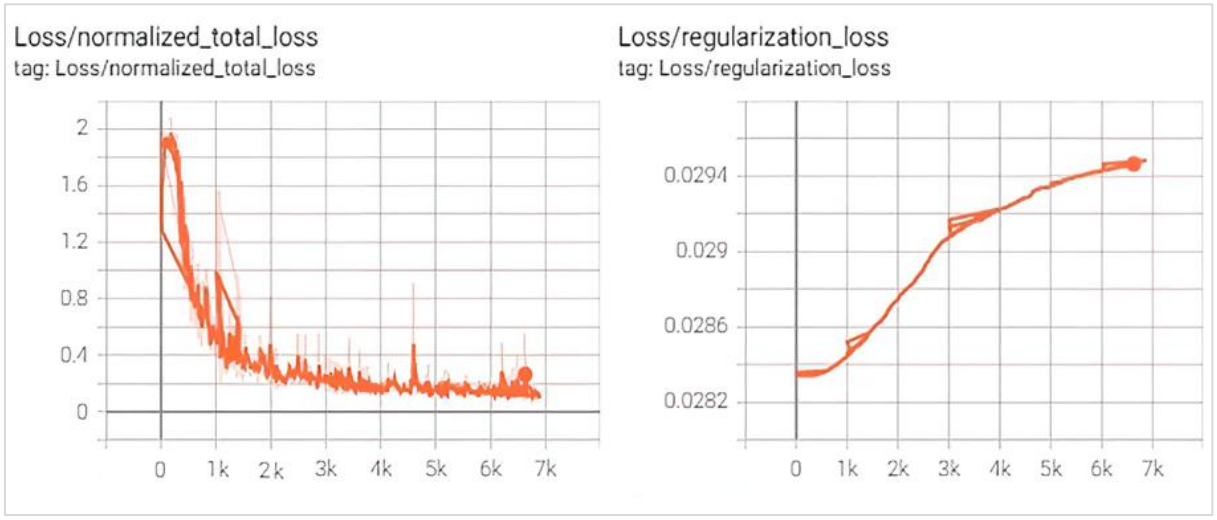


Figure 9 Normalized and regularization losses

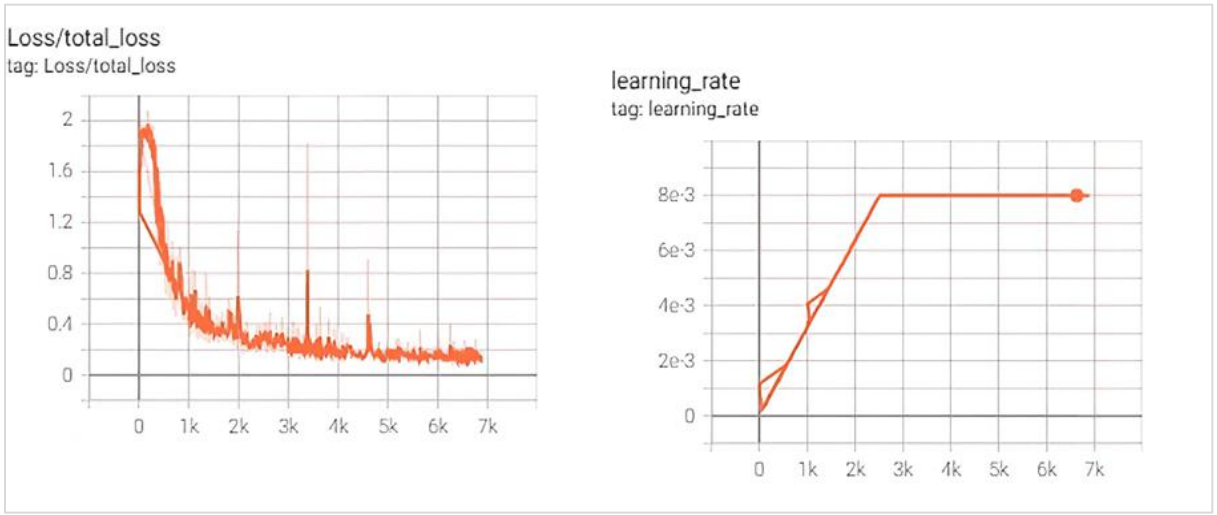


Figure 10 Total loss and learning rate

Table 1 illustrates the complete performance for the different models. Precision, recall, mAP, F1-score were calculated for different models. The mean Average Precision calculates a score by relating the observed box to the ground-truth bounding box. The greater the score, the more precise the model's object detection detections are. The weighted average of precision and recall yields the F1-score. The F1-score is a number between 0 and 1, with 1 indicating the highest level of accuracy.

The precision recall curve (P-R Curve) obtained with the Mask R-CNNs model is revealed in *Figure 13*. The average precision of the Mask R-CNN algorithm for gallstone detection is calculated by integrating the region under the P-R curve.

Table 2 depicts the implementation of Mask R-CNN with diverse foundation networks for Gallstone Object detection. It is clear that Mask R-CNN with backbone network of Resnet-101-FPN combination gives a very good result for the object detection. The Mask-RCNN model is good compared to SSD and Faster R-CNN for object detection of gall stones. *Figure 14* illustrates the object detection of gallstones by the three algorithms used.

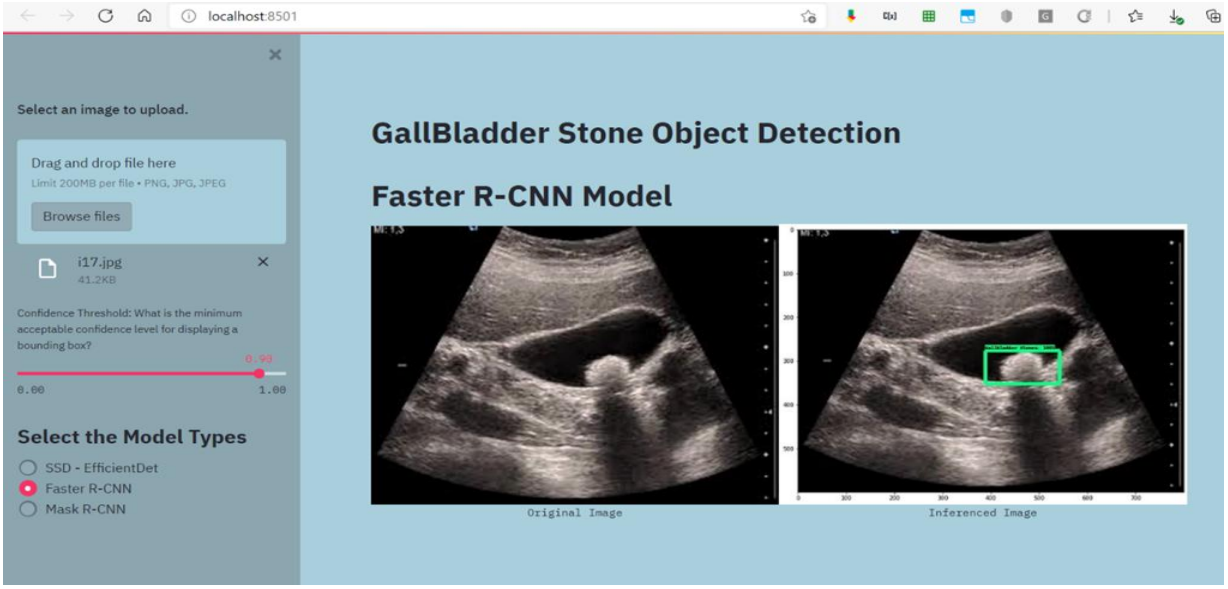


Figure 11 Results of faster R-CNN model from the healthcare informatics system

Figure 12 Results of Mask R-CNN from the healthcare informatics system

Table 1 Performance overview of the models

Parameters	SSD-efficientDet	Faster R-CNN	Mask-RCNN
Precision	0.6617	0.867	0.78
Recall	0.9325	0.945	0.986
mAP	0.567	0.67	0.77
F1-Score	0.79325	0.85	0.938

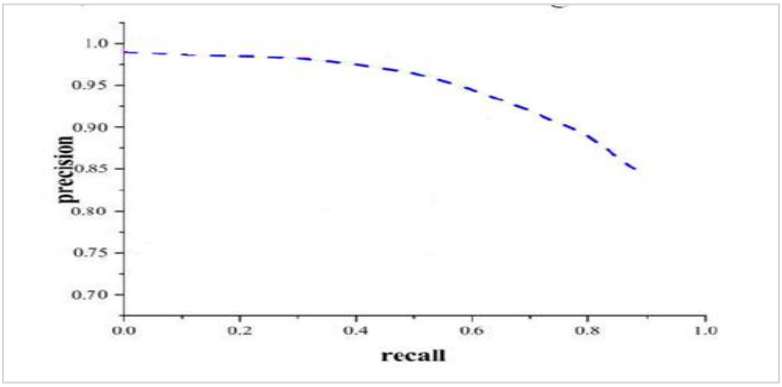


Figure 13 P-R curve

Table 2 Implementation of Mask R-CNN with backbone Network

Deep learning model	Backbone network	Mean average precision
Mask R-CNN	ResNet-101-C4	0.778
Mask R-CNN	ResNet-101-Feature Pyramid Network	0.789
Mask R-CNN	ResNeXt-101-Feature Pyramid Network	0.765

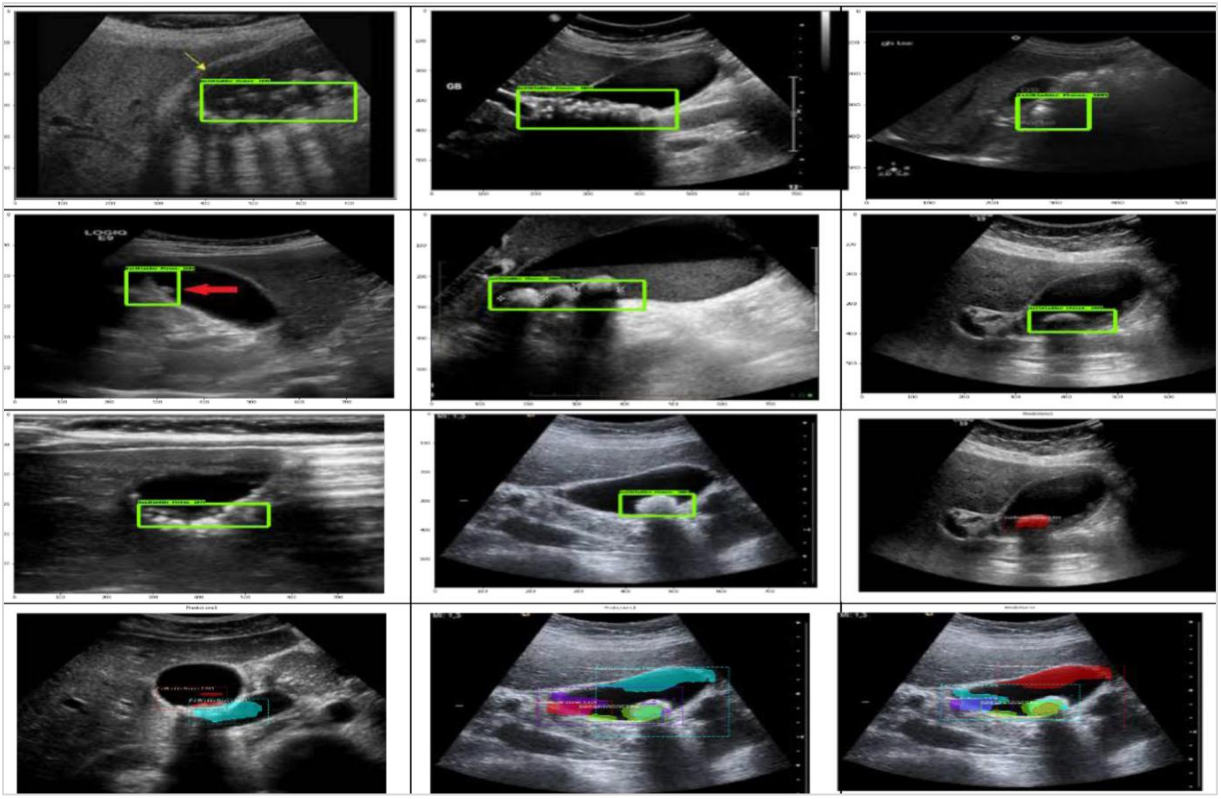
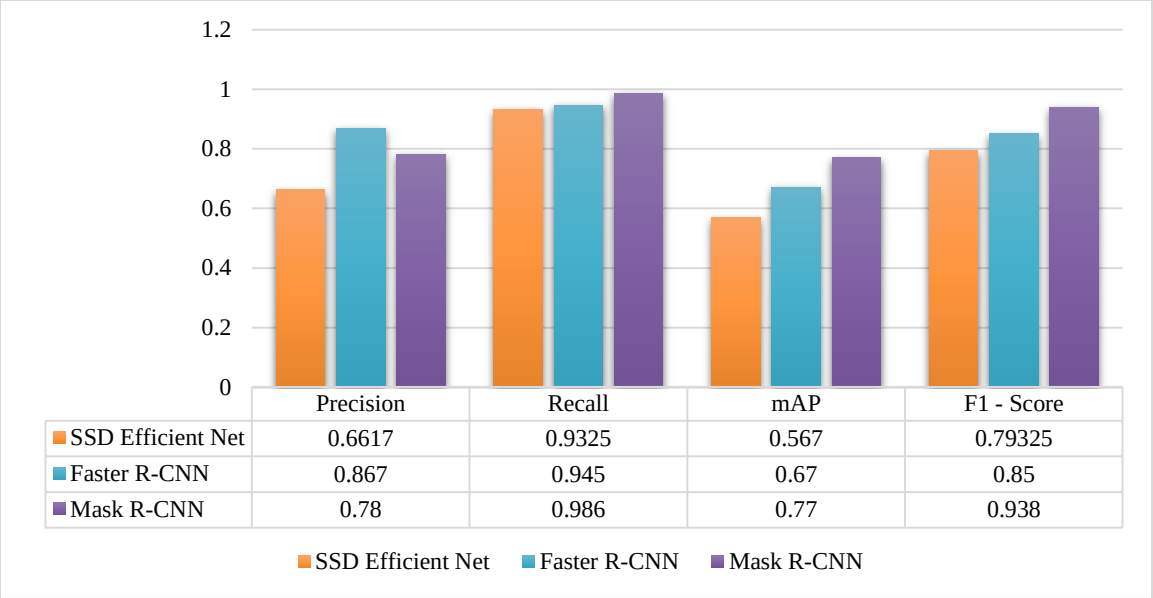


Figure 14 Displays the detection of the gallstones using the state-of-the-art models

5.Discussion

DL has revolutionized a number of domains in recent years, spurring breakthroughs in computer vision, natural language processing, and other areas. We envisioned to demonstrate the possibilities of this strong method when applied to CT images in this research. The models such as SSD, Faster R-CNN and Mask R-CNN has been used on the gallbladder CT images. The CT images are annotated using the VGG annotator and labelimg tools. Annotating requires the help of a medical practitoner with the domain knowledge. The models are compared and contratsed on some of the metrics. Precision, recall, mAP, and F1-score are measures used to show how effectively various models perform. Precision is the ratio of correctly classifying the presence of gallstones to all positively anticipated examples. Faster R-CNN is an excellent model for accurately detecting the occurrence of gallstones, in contrast to SSD and Mask R-CNN. Recall measures how many positive cases the classifier has classified as positive. It should be higher value. *Figure 15* illustrates the graphical depiction and comparison of machine learning techniques.

Mask R-CNN is having recall value of 0.986 compared to other two algorithms. F1−Score is the weighted average of the recall and precision. Mask R-CNN has a value of 0.938 which indicates that the classifier classifies it correctly. Manually annotating the images is the limitation. In further work we try to show the size of the gallstones, the calcification stages of the gallstone and whether the stone is in slurry state or in stone form. Getting more images and labelling is a challenging task. In the future, we would like to classify the stones into single stones, multiple stones, and stones in a slurry state.



	Precision	Recall	mAP	F1 - Score
SSD Efficient Net	0.6617	0.9325	0.567	0.79325
Faster R-CNN	0.867	0.945	0.67	0.85
Mask R-CNN	0.78	0.986	0.77	0.938

Figure 15 Result comparison of various machine learning techniques

6.Conclusion

Gallstones is a highly prevalent and severe disease and one of the main factors that causes hospitalizations around the globe. This DL based healthcare informatics system can be used as a supplement to bridge the gap among reviewers and to diminish the incidence of false positives. The data is collected from both rural and urban areas. The data is annotated with the help of medical professionals. We have demonstrated the training and testing of SSD-efficientDet, Faster R-CNN and Mask R-CNNs models to detect the gallstones. For object detection, these models were given an image, a model name, and a confidence score. The various parameters of performance evolution have been determined and tabulated. The Mask R-CNN model was modelled and trained using different backbone network. The results demonstrate the effectiveness of the approach.

Acknowledgment

None.

Conflicts of interest

The authors have no conflicts of interest to declare.

Author's contribution statement

Veena A: Data collection, Study conception, writing – review and editing, draft manuscript preparation. **Gowrishankar S:** Conceptualization, draft manuscript preparation, writing – original draft, writing – review and editing, analysis and interpretation of results, guidance.

References

[1] Yildirim M, Kurum H. Influence of poles embrace on in-wheel switched reluctance motor design. In 18th international power electronics and motion control conference 2018 (pp. 562-7). IEEE.

[2] Sun H, Tang H, Jiang S, Zeng L, Chen EQ, Zhou TY, et al. Gender and metabolic differences of gallstone diseases. World Journal of Gastroenterology: WJG. 2009; 15(15):1886-91.

[3] Shaffer EA. Epidemiology and risk factors for gallstone disease: has the paradigm changed in the 21st century? Current Gastroenterology Reports. 2005; 7(2):132-40.

[4] Portincasa P, Moschetta A, Palasciano G. Cholesterol gallstone disease. The Lancet. 2006; 368(9531):230-9.

[5] Heaton KW, Braddon FE, Mountford RA, Hughes AO, Emmett PM. Symptomatic and silent gall stones in the community. Gut. 1991; 32(3):316-20.

[6] Chen CH, Lin CL, Kao CH. Erectile dysfunction in men with gallbladder stone disease: a nationwide population-based study. American Journal of Men's Health. 2019; 13(2).

[7] Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current applications and future impact of machine learning in radiology. Radiology. 2018; 288(2):318-28.

[8] Van SRJ, Cohen R, Eldar YC. Deep learning in ultrasound imaging. Proceedings of the IEEE. 2019; 108(1):11-29.

[9] Van SRJ, Demi L. Localizing B-lines in lung ultrasonography by weakly supervised deep learning, in-vivo results. IEEE Journal of Biomedical and Health Informatics. 2019; 24(4):957-64.

[10] Anantharaman R, Velazquez M, Lee Y. Utilizing mask R-CNN for detection and segmentation of oral diseases. In international conference on bioinformatics and biomedicine 2018 (pp. 2197-204). IEEE.

[11] Haq MN, Irtaza A, Nida N, Shah MA, Zubair L. Liver tumor segmentation using resnet based mask-R-CNN. In international bhurban conference on applied sciences and technologies 2021 (pp. 276-81). IEEE.

[12] Hsieh YC, Chin CL, Wei CS, Chen IM, Yeh PY, Tseng RJ. Combining VGG16, mask R-CNN and inception V3 to identify the benign and malignant of breast microcalcification clusters. In international conference on fuzzy theory and its applications (iFUZZY) 2020 (pp. 1-4). IEEE.

[13] Sha G, Wu J, Yu B. Detection of spinal fracture lesions based on SSD. In proceedings of the 2020 international conference on aviation safety and information technology 2020 (pp. 539-42).

[14] Sokolova N, Taschwer M, Sarny S, Putzgruber-adamitsch D, Schoeffmann K. Pixel-based iris and pupil segmentation in cataract surgery videos using mask R-CNN. In IEEE 17th international symposium on biomedical imaging workshops (ISBI Workshops) 2020 (pp. 1-4). IEEE.

[15] Kim JU, Kwon J, Kim HG, Lee H, Ro YM. Object bounding box-critic networks for occlusion-robust object detection in road scene. In 25th international conference on image processing 2018 (pp. 1313-7). IEEE.

[16] Chen CW, Ruan SJ, Lin CH, Hung CC. Performance evaluation of edge computing-based deep learning object detection. In proceedings of the 2018 VII international conference on network, communication and computing 2018 (pp. 40-3).

[17] Chen L, Xie T, Wang X, Wang C. Identifying urban villages from city-wide satellite imagery leveraging mask R-CNN. In adjunct proceedings of the ACM international joint conference on pervasive and ubiquitous computing and proceedings of the ACM international symposium on wearable computers 2019 (pp. 29-32).

[18] Lin M, Chen C, Lai C. Object detection algorithm based AdaBoost residual correction fast R-CNN on network. In proceedings of the 3rd international conference on deep learning technologies 2019 (pp. 42-6).

[19] Mohedano E, Mcguinness K, Healy G, O'connor NE, Smeaton AF, Salvador A, et al. Exploring EEG for object detection and retrieval. In proceedings of the 5th ACM on international conference on multimedia retrieval 2015 (pp. 591-4).

[20] Kanakatte A, Ramaswamy A, Gubbi J, Ghose A, Purushothaman B. Surgical tool segmentation and localization using spatio-temporal deep network. In 42nd annual international conference of the IEEE engineering in medicine & biology society 2020 (pp. 1658-61). IEEE.

[21] Jain R, Sutradhar A, Dash AK, Das S. Automatic multi-organ segmentation on abdominal CT scans using deep U-Net model. In OITS international conference on information technology 2021 (pp. 48-53). IEEE.

[22] Obaid AM, Turki A, Bellaaj H, Ksontini M. Detection of biliary artesia using sonographic gallbladder images with the help of deep learning approaches. In 8th international conference on control, decision and information technologies 2022 (pp. 705-11). IEEE.

[23] Qifang X, Guoqing Y, Pin L. Aircraft detection of high-resolution remote sensing image based on faster R-CNN model and SSD model. In proceedings of the international conference on image and graphics processing 2018 (pp. 133-7).

[24] Basu S, Gupta M, Rana P, Gupta P, Arora C. Surpassing the human accuracy: detecting gallbladder cancer from USG images with curriculum learning. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2022 (pp. 20886-96).

[25] Tao Y, Zongyang Z, Jun Z, Xinghua C, Fuqiang Z. Low-altitude small-sized object detection using lightweight feature-enhanced convolutional neural

network. Journal of Systems Engineering and Electronics. 2021; 32(4):841-53.
[26] Song Z, Zhang Y, Liu Y, Yang K, Sun M. MSFYOLO: feature fusion-based detection for small objects. IEEE Latin America Transactions. 2022; 20(5):823-30.
[27] Liu L, Fan J, Xu C. Enhanced small object detection neural network. In proceedings of the international conference on aviation safety and information technology 2020 (pp. 665-9).
[28] Rizos P, Kalogeraki V. Deep learning for underwater object detection. In 24th pan-hellenic conference on informatics 2020 (pp. 175-7).
[29] Jung D, Shim S, Choo C, Hwang D, Nah Y, Oh S. A preliminary result of food object detection using swin transformer. In proceedings of the 8th international conference on computer technology applications 2022 (pp. 183-7).
[30] Xu Z. High-resolution object detection with SAR imaging for autonomous driving. In 5th international conference on compute and data analysis 2021 (pp. 91-5).
[31] Park H, Park S, Joo Y. Detection of abandoned and stolen objects based on dual background model and mask R-CNN. IEEE Access. 2020; 8:80010-9.
[32] Reddy SP, Kandasamy G. Cusp pixel labelling model for objects outline using R-CNN. IEEE Access. 2021; 10:8883-90.
[33] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: single shot multibox detector. In European conference on computer vision 2016 (pp. 21-37). Springer, Cham.
[34] 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.
[35] He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In proceedings of the IEEE international conference on computer vision 2017 (pp. 2961-9).
[36] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 779-88).
[37] Thara DK, Premasudha BG. A review on computer aided diagnosis of epilepsy using machine learning and deep learning. International Journal of Research and Analytical Reviews. 2018; 5(3).
[38] Erhan D, Szegedy C, Toshev A, Anguelov D. Scalable object detection using deep neural networks. In proceedings of the IEEE conference on computer vision and pattern recognition 2014 (pp. 2147-54).
[39] Szegedy C, Reed S, Erhan D, Anguelov D, Ioffe S. Scalable, high-quality object detection. arXiv preprint arXiv:1412.1441. 2014.
[40] Girshick R. Fast R-CNN. In proceedings of the IEEE international conference on computer vision 2015 (pp. 1440-8).
[41] Song F, Wu L, Zheng G, He X, Wu G, Zhong Y. Multisize plate detection algorithm based on improved mask R-CNN. In international conference on smart internet of things (SmartIoT) 2020 (pp. 277-81). IEEE.
[42] Thara DK, Premasudha BG, Xiong F. Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. Pattern Recognition Letters. 2019; 128:544-50.
[43] Parihar J, Kansal P, Singh K, Dhiman H. Assessment of bioinformatics and healthcare informatics. In Amity international conference on artificial intelligence 2019 (pp. 465-7). IEEE.
[44] https://docs.python.org/3/library/pickle.html. Accessed 03 October 2022.
[45] https://github.com/tensorflow/models/tree/master/research/object_detection. Accessed 03 October 2022.
[46] https://www.tensorflow.org/tensorboard. Accessed 03 October 2022.

Veena A is currently pursuing her Ph.D. in Computer Science and Engineering from Visvesvaraya Technological University, Belagavi, India. She earned her M.Tech. in Computer Science and Engineering and B.E. in Information Science and Engineering from Visvesvaraya Technological University (VTU), Belagavi, India in the years 2011 and 2004 respectively. She has published papers in various reputed International Journals and Conferences. She has co-authored a book and has earned a patent. Her research interests include Automated Machine Learning, Deep Learning, Image Processing, Natural Language Processing, and Reinforcement Learning in the Smart Healthcare Domain. She has been serving as an Assistant Professor in the Department of Computer Science and Engineering at Dr. Ambedkar Institute of Technology, Bengaluru, India.
Email: veenaa1@acm.org

Gowrishankar S is currently working as a Professor in the Department of Computer Science and Engineering at Dr.Ambedkar Institute of Technology, Bengaluru, India. He earned his M.B.A. in Marketing Management from Indira Gandhi National Open University (IGNOU), New Delhi, India in 2020, Ph.D. in Engineering from Jadavpur University, Kolkata, India in 2010, M.Tech. in Software Engineering from Visvesvaraya Technological University (VTU), Belagavi, India in 2005 and B.E. in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belagavi, India, in 2003. He has published several papers in various reputed International Journals and Conferences. He is serving as editor and reviewer for various prestigious International Journals. He has earned two patents. His current research interests are mainly focused on Data Science, including its technical aspects as well as its applications and implications.
Email: gowrishankarnath@acm.org

Appendix I

S. No.	Abbreviation	Description
1	2D	Two Dimensional
2	AI	Artificial Intelligence
3	API	Application Programming Interface
4	BMI	Body Mass Index
5	CNN	Convolutional Neural Network
6	COCO	Common Objects in Context
7	CT	Computerized Tomography
8	DL	Deep Learning
9	EEG	Electroencephalogram
10	EOSD	Enhanced Small Object Detection Neural Network
11	Faster R-CNN	Faster Region- based Convolutional Neural Network
12	FCN	Fully Convolutional Network
13	GSD	Gallstone Disease
14	IoT	Internet of Things
15	IoU	Intersection over Union
16	mAP	Mean Average Precision
17	MC	Micro Calcification
18	ML	Machine Learning
19	Mask R-CNN	Mask Regional-Convolutional Neural Network
20	NCS	Neural Compute Stick
21	OBB	Object Bounding Box
22	P -R curve	Precision Recall Curve
23	RoI	Region of Interest pooling
24	RPN	Region Proposal Network
25	SSD	Single Shot Detector
26	VGG	Visual Geometry Group
27	YOLO	You Only Live Once