

A Deep Learning Approach to Detect Endometrial Tuberculosis

Varsha Garg; Vikas Saxena; Anita Sahoo

{[varsha.garg](mailto:varsha.garg@jiit.ac.in), [vikas.saxena](mailto:vikas.saxena@jiit.ac.in), [anita.sahoo](mailto:anita.sahoo@jiit.ac.in)}@jiit.ac.in

Jaypee Institute of Information Technology, A-10, Sector-62, Noida-201309,
Uttar Pradesh, India.

ABSTRACT

Purpose:Female Genital Tuberculosis (FGTB) is found to be a major reason for infertility; but is currently an under-researched medical condition. In order to identify the cause of infertility, women have to undergo several invasive and expensive tests. Endometrial Tuberculosis accounts for 70% of infertile FGTB infected patients. Since, Transvaginal Ultrasound is the initial non-invasive investigational mode for infertility assessment; this research is focused on identification of endometrial tuberculosis.

Originality:This is a novel effort towards developing a computational method to identify Endometrial Tuberculosis within Transvaginal (TVS) ultrasound images in order to assist clinicians in identifying infertility without delay.

Methodology:In consultation with medical experts, real time TVS Ultrasound images have been collected from a leading Hospital in Delhi, India. Different augmentation techniques suitable for ultrasound imaging have been applied on images of 67 patients to increase the training data size. Experiments have been conducted implementing different Deep Neural Network models.

Findings:Among all the tested models, the layered architecture of Convolution Neural Networks, Region Propagation Network and ResNet50 was found to be more effective in identifying TB in the endometrium. Further, images from 17 patients were collected and used to test the model, which shows an average predictive accuracy of 84.9% and a F1-measure score of 0.851. The model also recorded a sensitivity of 88%.

KEYWORDS: Female Genital Tuberculosis; Transvaginal Ultrasound Image Analysis; Deep Learning. Convolution Neural Networks; Endometrial Tuberculosis; Machine Learning.

1. INTRODUCTION

India is in a “Catch-22” situation as the fertility rate is on a dramatic decline as per SRS statistical Report 2016 presented in National Health Policy, 2018. According to Dr. Sama Bhargava, an IVF expert and Consultant, Fortis Hospital, Noida, in her article in Diplomat dated 30th May 2018 “Infertility is an under-researched condition that is wrecking marriages and even people’s lives”. Infertility patients are exposed to invasive and expensive tests at the private centres without preliminary normal investigations. Unexplained Infertility is then further investigated making it dreadful for the patient.

“GTB lying dormant in the uterus decreases anti-Mullerian hormone (AMH) levels by 30% leading to poor egg quality and count leading to infertility establishing as the reason of failed in-vitro fertilization”, as per IVF specialist, Dr. Padma Rekha Jirge (Jirge, 2016). The biggest presentation of FGTB is Infertility (Jahromi, et al., 2001; Sharma, 2015). AIIMS recorded 26% prevalence of FGTB in infertile women and an incidence of 42.5% cases of infertility in FGTB (Sharma et al. 2016). About 25% of Indian women seeking Assisted Reproduction Techniques suffered from GTB and about 50% women with tubal infertility had GTB and a large

number of women earlier diagnosed with TB approached the doctor for infertility (Grace et al., 2017; Legro, R. S. et al., 2016). Hence, the gravity of GTB on infertility can be understood. Diagnostic technique of GTB should be highly sensitive, quick and reliable in its early stage leaving a possibility of cure before the tubes are damaged beyond recovery. These are generally the endometrium in 70% FGTB cases and 10-15% ovarian involvement in a TVS ultrasound image (Sharma et al., 2016). As per the WHO, diagnosis of EPTB should be made on the basis of 'one culture-positive specimen, or positive histology or strong clinical evidence consistent with active EPTB' (WHO, 2016).

FGTB diagnosis involves presence of unexplained infertility with other pelvic symptoms, past history of personal or contact TB and an abnormal chest X-ray. Probable diagnosis using Ultrasound abnormalities include endometrial complications of thinning, hyper-echogenicity and distortion or presence of endometrial fluid could initiate an anti-tubercular treatment. With inconclusive findings, further investigations for histopathological and microbiological evidence with endometrial biopsy were suggested. Further, hysteroscopy, laparoscopy or surgery obtained the tissue and an AFB stain could confirm the diagnosis.

Recently, Transvaginal (TVS) Ultrasound has become the most powerful tool for the investigation and management of female infertility and is well accepted. When a patient approaches a specialist for Infertility, a Transvaginal Ultrasound is always done for initial investigation. During this if the patient is suffering from FGTB it should be diagnosed before any other invasive methods are used to further investigate the cause of Infertility. But, manual image interpretation, human limitations and large inter-grader variability in medical diagnosis are the major roadblocks. The limitation is the subjectivity that can also result in erroneous reporting reducing benefits. Training and gaining experience to be an excellent radiologist is long term, thus a computational method becomes a powerful assisting tool for objective diagnosis. Computer assisted research is required within developing countries where the global burden of infertility is the greatest, requiring greater emphasis on innovative, fast, safe and cost-effective solutions to sub-fertility / infertility diagnosis.

Although such methods exploiting ultrasound images exist, but those are found suitable for thyroid nodule diagnosis (Nugroho et al., 2019), carotid image classification, the breast and liver cancer, the gastroenteric and cardiovascular diseases, spine curvature, and the muscle disease (Shung, 2011); there is no such effort made for developing computational methods for assisting in diagnosis of FGTB from Ultrasound images. In recent past, machine learning algorithms have been successfully used to develop these methods for other medical issues. The efficiency of such models depends on presentation of meaningful data for training. Since, Ultrasound images are inherently ill-defined; therefore it becomes difficult to extract meaningful features for differentiating between normal and abnormal patterns. Although machine learning approaches have been used very effectively for diagnosis using computational expertise; feature extraction still remains a big challenge. Deep Learning models actually overcome this problem of segmentation and handcrafted feature extraction (Liu et al., 2019). Using deep learning models automatic features are extracted outperforming the traditional classification (Khan, et al., 2019). Recently Deep Learning in medical imaging is effectively used for detection of Cancers (Yassin et al., 2018), Diabetic Retinopathy (Mansour, 2018), Histological and Microscopical Elements, Gastrointestinal (GI) Diseases, Cardiac Imaging, Tumour, Alzheimer's and Parkinson's Diseases. Very recently it has successfully predicted Alzheimer's six years before the onset (Razavi, et al., 2019). GoogLeNet was used for breast lesion detection with 90%

accuracy (Han et al., 2017). VGGNet and fully connected network was used for liver lesions with 93% accuracy ((Meng et al., 2017) whereas feature extraction using CNN and SVM classification achieved 96.8% accuracy (Liu et al, 2017; Huang et al., 2018). CNN was used for feature extraction and LSTM for classifying Fetal Ultrasound Standard Plane detection with 86% accuracy (Chen et al., 2015). For thyroid nodules, GoogLeNet reported 99.13% accuracy (Chi et al., 2017).

In this paper focus is on identification of endometrial tuberculosis within TVS ultrasound images. Various CNN models have been used for the identification task. A brief study of these learning models in image analysis, issues and purpose of evolution are discussed in the section 2. The detailed approach is explained in section 3. The medical data collection and augmentation using different augmentation methods suitable methods for ultrasound images is described in section 3.1. The proposed algorithm and the experimental results are discussed in section 3.2 and section 4 respectively.

2. CONVOLUTION NEURAL NETWORKS IN IMAGE ANALYSIS AND RELATED ISSUES

Convolutional Neural Network (CNN) (Krizhevsky et al., 2012) is a Deep Neural Network for the image classification, object detection and segmentation with strong neighbour data correlations. A typical CNN architecture has a series of convolution and pooling layers followed finally by multiple fully connected layers. The convolution layers apply filters acting as feature extractors producing feature maps as input at each layer. Pooling layer then increases robustness to spatial variations (feature shifting or scaling) along with dimensionality reduction. AlexNet (Krizhevsky et al., 2012), the most well-known, general classification CNN architecture outperforms the fully connected neural networks and the existing machine learning methodologies in image classification on ImageNet. AlexNet has five convolution layers, three pooling layers, and two fully-connected layers with approximately 60 million free parameters.

More complex CNN models have larger number of parameters and can learn both local and global structures in images performing very well in CAD classification problems. Deeper architectures like GoogLeNet [21 layers] and VGG [13-16 layers] emerged later in 2014 and 2015 respectively (Szegedy et al., 2014; Simonyan and Zisserman, 2015). VGG is a simple model replacing large (11×11 and 5×5) kernel-sized filters in the early convolutional layer of AlexNet with stacked 3×3 convolution and pooling layers followed by fully connected layers. Two stacked 3×3 convolution layers create a receptive field similar to 5×5 . And 3 stacked 3×3 produce a receptive field similar to 7×7 . Multiple ReLU functions also introduce more non-linearity.

However, due to exploding and vanishing gradients, there is a loss of generalization after a certain depth (Simonyan and Zisserman, 2015). Therefore, there is a need to identify sufficient number of layers that could handle the non-linearity in the given classification problem. This problem of training deeper networks has been reduced by introducing a new neural network layer, the residual block. These residual connections make it easier and faster to train as compared to other very deep models; VGGNet and GoogLeNet. This improves on learning capability for classification and object detection (He et al., 2016). Also they introduce an “identity shortcut” connection skipping one or more layers making them more efficient than the stacked models. ResNet50 surpassed human performance on the ImageNet dataset (He et al., 2015; Alom, et al., 2018) as the gradients could flow directly through skip connections backwards from later

layers to initial filters. Early layers of deeper models can be replaced with a shallow network and remaining layers are identity functions. Each identity block in ResNet50 is 3-layers deep. This enhances training speed of the deep networks and reduces number of parameters by increasing the network's depth instead of width. It achieves higher accuracy in Image Classification (Grossman, 2019).

Further, using Region Propagation Network for Object Detection (Ren et al., 2016) increased recognitions abilities of CNNs. The feature maps from the CNN model serves as input to a RPN returning an object proposal or Region of interest, with a score, object-ness score denoting the presence and absence of an object in the region. Anchors are bounding boxes are rectangular but with different size and aspect ratio. A feature map has 3 dimensions; height, width and depth, so anchor has the same height and width. RPN takes anchors separated by a stride given by r pixels. It scans over the feature maps by sliding the anchor boxes of multiple scales and aspect ratio calculating two scores indicating the presence of an object and the bounding box regression which could be used to fit to the actual object better during CNN processing. The convolution feature maps from the CNN is passed through a 3×3 filter. Further they are passed through two parallel 1×1 filters with number of channels being channel dependent. Redundant regions are obtained by RPN as multiple regions are proposed for the same object. Then non-maximum suppression based on the class score narrows down the number of regions of interest (ROIs). ROI's are converted into fixed-sized vectors through ROI pooling and input to a fully connected layer. Bounding box regression layer outputs 4 parameters (coordinates of centre pixel, width and height or (x,y) coordinates of two diagonally opposite points of a box) for the bounding box whereas the classification layer gives the object-ness score of the box (for both background and foreground). It is conducted in a supervised manner by fully connected layer.

In medical research as the availability of labelled data is unfortunately scarce, therefore Transfer Learning is preferred (Pan et al., 2009; Tan et al., 2018). It is a machine learning method used in predictive modelling when a model developed for one classification task is used as the initial starting point for the other classification task resulting in training speedup and also improves the performance of the deep learning model. Training a deep learning model from scratch is very expensive therefore transfer learning uses pre-trained weights for initial layers and then fine tunes for further layers.

In this paper, a deep learning computational model of identifying endometrial tuberculosis underlying female genital tuberculosis in TVS ultrasound images of infertility patients is being presented. The aim is to here isto identifyEndometrial TB by recognizing, classifying and localizingpotential regions of interestwithin the presented ultrasound image. We explore the use of TVS Ultrasound in the disease as it is non- invasive and easy to perform. Due to lack of prior knowledge for extracting suitable features from ultrasound images to characterize Endometrial TB, it is quite a challenging problem in ill-defined ultrasound images. Therefore, DNN framework is used for the problem in hand. For improved performance, current focus of the researchers in the field of DNN are on implementation of larger networks like ResNet, VGGNet etc. or use of multiple networks, where after independent processing the responses are combined. But, there are problems for which simple DNN such as AlexNet is also performing well and it doesn't require using a complex network structure. Here, aim wasto find a suitable DNN model for Endometrial TB classification from ultrasound images. Strengths of different deep neural networks have been explored to design a layered approach for automated diagnosis of Endometrial TB in TVS ultrasound images. The DNN architecture inspired by faster RCNN (Ren

et al., 2015) has shown the better performance for the dataset in hand. For small datasets, DNNs don't perform well; therefore, suitable augmentation techniques have been applied to increase the data size.

3. PROPOSED METHOD

The process of identifying Endometrial TB in the ultrasound image involves four main tasks; Data preprocessing, feature extraction, potential region proposal and finally classification of the presented image as Normal and Abnormal. The preprocessing task involves data augmentation to increase the dataset size. The augmentation methods are chosen in order to make the method rotation, translation and contrast invariant. Feature extraction, region proposal and classification are implemented using Convolutional Neural Network. The detailed methodology is explained in the following sections.

3.1 Data Collection and Augmentation

For identification of infertility due to Endometrial TB, USG images of patients coming to Sai Clinic, Delhi, India is collected under the supervision of Dr. Vibha Bansal. For the dataset shared, images of 67 patients [31 normal and 36 abnormal] have been acquired using Voluson P8 Ultrasound machine. TVS USG imaging modality has been used and they are labeled using her expertise. During the period of our study (September 2016–February 2019), the samples from infertile women visiting the gynecologists at two centers in Delhi and Ghaziabad were analyzed. The inclusion criteria was based on patient history of symptoms varying from unexplained infertility in age group of 18–40 years, absent or irregular menstruation cycle with scanty flow, pelvic pain, and other menstrual disorders like painful cramps, general weakness and at times excessive bleeding leading to the abortions. The exclusion criteria considered women above reproductive age of 40 years, pulmonary tuberculosis suggestive patients with normal examinations of both abdomen and vagina, pregnant and nursing women.

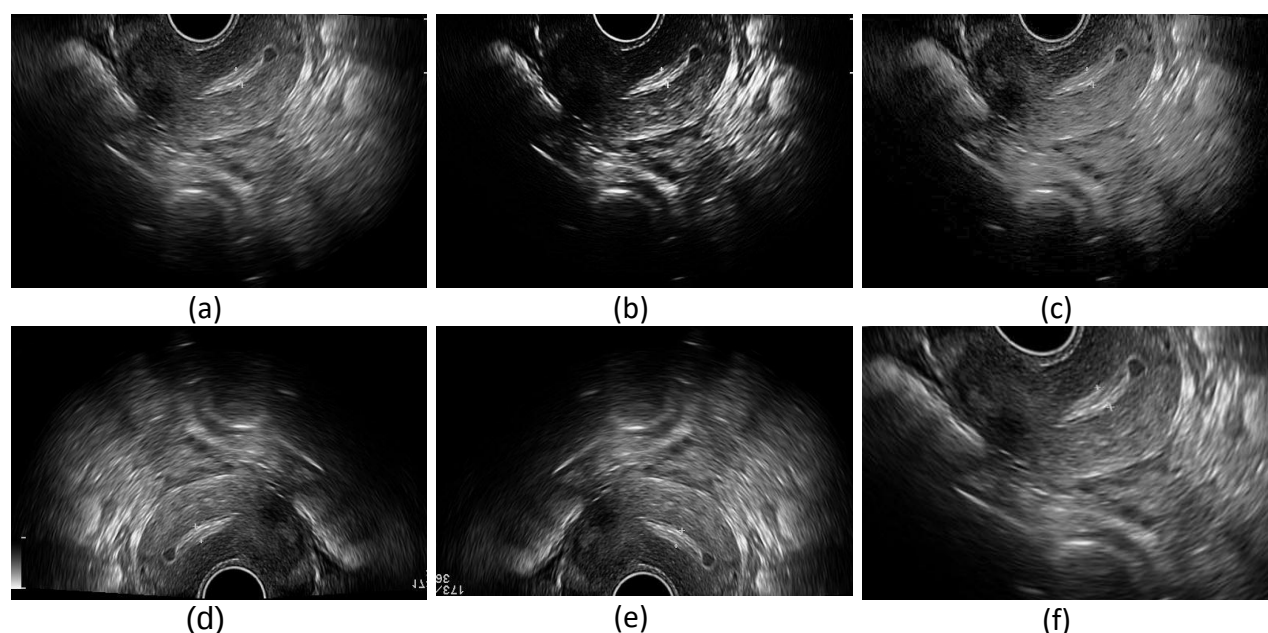


Fig. 1 Sample Dataset Images (a) Original image (b)-(f) Sample augmented images

Limited annotated datasets pose a challenge for CNN learning so data augmentation techniques which are meaningful for ultrasound imaging have been used. This includes five random repetitions of each of following techniques; rotation of -20 degree to +20 degree with random left right flip and translation of 25-50 pixels in all the four directions. Speckle noise is an inherent property of ultrasound imaging that normally reduces contrast in the image. Another augmentation method of adding contrast ranging from -4 to +4 intensity values is used to generate 2 additional images per patient. This increases dataset size by twelve times to generate a total of 871 images. Using these techniques reduces the problem of over-fitting in CNN due to lesser number of images originally available. The dataset now contains 871 annotated image slices and 2 labels, namely normal and abnormal. The images were finally resized to 224×224 . Original image sample is shown below in Fig. 1(a) and few augmented images are shown in Fig. 1(b)-(f).

The bounding box and labeling procedure for the dataset was done by the expert. These bounding boxes were actually the Region of Interest (ROI), expert deems important for the diagnosis. These bounded and labeled images are used as ground truths for training and validating the model. To create a bias free dataset, stratified random sampling was done and data was divided into training and validation in the ratio of 70:30. For testing the model's prediction ability, 17 new images (7 normal and 10 abnormal) were collected from new patients.

3.2 Deep Neural Network based Endometrial Tuberculosis detection

The images in the training dataset prepared are presented to a pre-trained CNN network of ResNet50. It is used as the backbone network to extract the feature maps of the image (F). These 1024 feature maps are extracted at activations_40_ReLU layer. Then these maps are passed to a Region Proposal Network (RPN) and object proposals thus are obtained by training the bounding box regression and classification layer minimizing the losses. These Proposals (RP) and the original Feature maps (F) now pass to the residual convolution layers of the backbone network extracting advanced feature maps of only the proposed regions given by RPN by sending the boundary coordinates and objectness of the bounding box. These are then forwarded to ROI pooling layer to return a fixed size feature map. Then, fully connected layers use Softmax for final object classification and bounding boxes prediction. The algorithm for the above proposed approach is described as below.

Input: Transvaginal Ultrasound images $I_1, I_2, I_3, I_4, I_5, \dots, I_N$.

Output: Classified Image with a bounding box around the Region of Interest (Abnormal or Normal) with a probability /confidence score associated with each bounding box and class label. Accuracy is returned based on correct classification with the model.

Process:

- 1) For each training image, I_i (where $i = 1$ to N)
- 2) I_i = read an image from the folder
- 3) Net = ResNet50 /* Load pre-trained network (Transfer Learning).
- 4) Layer = activations_40_ReLU /* Extract the feature map F at current layer
- 5) F = ResNet50(I_i , Layer) /* Feature map of size [14,14,1024] returned
- 6) Do 7) till a specified number of epochs or $\text{loss} < 0.05$
- 7) $RPI[\text{objectness_score}, \text{bounding_boxDeltas}] = \text{Region_Proposal_Network}(F)$ /* Connect the Feature extraction layer above with RPNetwork. (3 × 3 convolution layer and then to 1 × 1 Classification and regression Layers.

- 8) $New_Layer = conv_5_0$
- 9) $FP = ResNet50(New_Layer, F, RPi) /* Feature map of proposals$
- 10) $O_i = ROI_Pooling(FP) /* ROI pooling layer gives fixed size features maps of regions.$
- 11) Do step 12) for epochs= 14K or $loss \leq 0.05 /* Total Loss, L = L_{CLS} + L_{BOX}$
- 12) $[classification_score, bounding_box_coordinates] = classify(Softmax, O_i)$

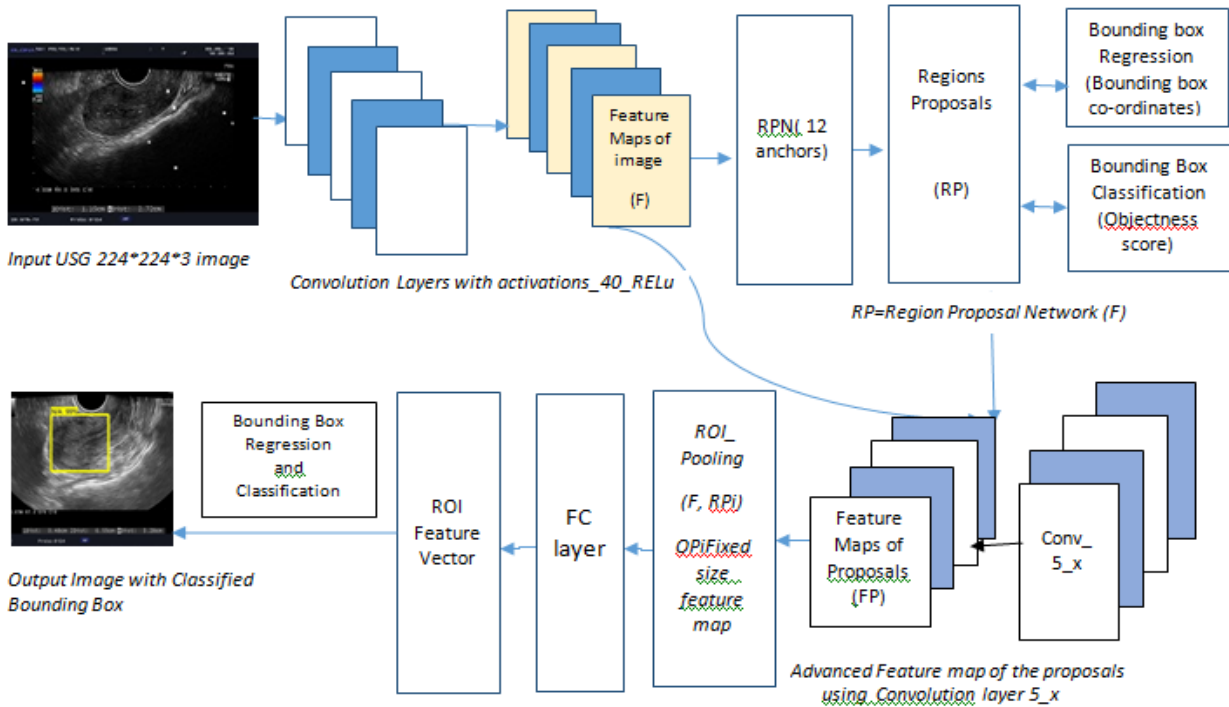


Fig. 2 Process of Endometrial Tuberculosis identification and classification

The pre-trained CNN network, ResNet50 is initialized for transfer learning to our dataset. The existing fully connected layers for classification are removed and a Region Proposal Network connected after the feature extraction layer (activations_40_ReLU layer). Images are then passed through the network and 1024 feature maps (F) of size 14×14 are returned at this point. Now the RP Network with a 3×3 convolution layer generates the region proposals which are fine-tuned using the two layers, a 1×1 Bounding Box Classification(convolution layer and again a 1×1 bounding Box regression convolution layer).The network is then fine-tuned for the given number of epochs so that the model converges to allowed loss. Region proposals (RP_i) with 4 co-ordinates for the bounding box dimension and an objectness score is returned in the structure. This structure in conjunction with the feature map (F) pass through the remaining convolutional layers for feature map of the proposals (FP). They are of varied anchor sizes. To make them of a fixed size (O_i), ROI_Pooling layer is used. Then a ROI_feature vector through a fully connected layer is generated for classification using Softmax. The classifier returns the bounding box coordinates with the confidence score for a class.

Once the model is trained, it can be presented with ultrasound image of a patient to identify Endometrial TB if any. The process is as shown in the Fig. 2 above.

4. IMPLEMENTATION, RESULTS AND ANALYSIS

Intel I7 4th generation CPU and NVIDIA 760M graphics processing unit has been used for implementation. The learning models are pretrained on ImageNet dataset. Here we used TensorFlow-slim model Library which is a deep learning framework with Python language. Caffe model was used for implementation of AlexNet.

Different deep neural networks such as Region Proposal Network with ResNet50, Convolution Neural Networks of three and five layers with SVM classifier, AlexNet with Softmax Layer and then AlexNet with SVM classifier (after extracting features at ReLU_5 after Convolution layer 5), CNN and RPN with VGG16 are implemented and experimented with the prepared dataset.

4.1 Model Configurations and Parameter Settings

The input layer for each network contains 50,176 neurons and output layer has 2 neurons. The configuration of the fully connected layers is the same in all networks. Rectified Linear Unit (ReLU) activation function have been used in all convolution layers and in all the networks as it trains faster by diminishing vanishing gradient problem. It is defined as given by equation 1.

$$f(x) = \max(0, x) \quad \dots\dots\dots (1)$$

The initial learning rate ($initial_{LR}$) for all the networks is initialized to 0.001; momentum to 0.9 and decay to 0.01. The learning rate for the current iteration, LR is given by equation 2.

$$LR = initial_{LR} * \left(\frac{1}{1 + decay * epochs} \right) \quad \dots\dots\dots (2)$$

The optimizer used in all the networks was Adam. The number of epochs is set to 14K and momentum was fixed to 0.9 experimentally. Horizontal and vertical stride has been set to 16. L2 regularizer (ridge loss) is used for localization.

CNN model with three layers and five layers have two and four convolution layers respectively with a filter size of 7×7 in the beginning and 3×3 in later layers.

AlexNet with Softmax Layer was composed of five convolutional layers followed by 3 fully connected layers. The filter size of first convolution layer was 11×11 followed by a 5×5 filter in the second layer. Rest of the layers have a size of 3×3 .

AlexNet with SVM classifier has five convolution layers and the three fully connected layers of AlexNet were replaced by the Linear SVM (Support Vector Machine) classifier.

VGG16 uses 3×3 filters with a small receptive field. The stride was fixed to 16 and the padding was 1 pixel for convolution layers. Five max-pooling layers of size 2×2 and a stride of 2 are used. Three Fully-Connected (FC) layers follow a stack of 13 convolutional layers and the final layer was the Softmax layer. Feature extraction was done at conv5_1 layer.

A ResNet50 network has 49 convolution layers and 1 fully connected layer. It was initialized with a 7×7 filter followed by a 3×3 max pooling layer and a stride of 2. Four more blocks of convolution layers comprising of 3 convolution layers of sizes 1×1 , 3×3 and 1×1 are then added with identity mapping as a memory function.

For the Region Proposal Networks, 12 anchors of scale [0.25, 0.50, 1.0, and 2.0] and aspect ratio [0.50, 1.0, and 2.0] are being used. This network was fully connected to a 3×3 spatial window of the input convolution feature map. Each sliding window was mapped to a lower-dimensional vector of 512-d. This was followed by two 1×1 convolution layers for the regressor and the classifier. For every window 12 regions are proposed with regressor giving the 4 coordinates for

the bounding box and classifier giving score for object-ness (presence of an object). Stride of 1 was used. Smooth L1 with smoothing parameter of 0.6 was used. A distance metric defined as Intersection over Union, (IoU) was set to be greater than or equal to 0.6 for foreground and less than or equal to 0.1 for background. Object detection system extracts a maximum of 300 region proposals using Non Maximal Suppression (NMS) technique.

4.1.1 Loss Functions

Both the models using CNN with RPN involve four different losses; classification and localization losses each for region proposal Network and classification network. Regression loss was calculated only for foreground anchors. Total Loss, L is the weighted sum of classification loss and bounding box (localization) loss in both proposal and detector networks as given in Equation 3:

$$L = w1 * L_{CLS} + w2 * L_{BOX} \dots\dots\dots (3)$$

The weight parameters set as $w1=0.6$; $w2=0.4$ in order to give more weightage to classification loss. Classification Loss is given in Equation 4, where $p(i)$ is the predicted bounding box and $a(i)$ is the actual ground truth box provided by the expert. Intersection between the two is used with an IoU metric and using NMS techniques and loss optimization, the boxes are finally proposed. The localization losses, L_{BOX} are given as L1 smooth loss between ground truth box and bounding box coordinates.

$$L_{CLS}(p(i), a(i)) = \sum_{i=0}^{i=n} -a(i) \log p(i) - (1 - a(i)) \log (1 - p(i)) \dots\dots\dots (4)$$

$$L_{BOX} = \sum_{i=0}^{i=n} a(i).L_{smooth}(coordinates_{predicted} - coordinates_{actual}) \dots\dots (5)$$

These losses are accumulated according to Equation 3 to find total losses for the proposed model.

4.2 Training and Validation

Pre-trained models were initialized and transfer learning was used. The weights obtained for earlier dataset were used to fine-tune the model at later layers for faster and better performance. The training procedure was done after feature extraction step. For all the network models, a 5-fold validation was performed.

CNN with 3 layers and SVM classifier had an accuracy of 66% whereas a CNN with 5 layers and SVM classifier had an accuracy of 68.01%. AlexNet with Softmax layer has an average accuracy of 68.4%. The Softmax layer when replaced with SVM results in an increment of 8% giving an average accuracy of 76.24%. This shows that the earlier layers contain generic features e.g. edge detectors or textural features that are useful for finding abnormality in ultrasound image. These networks only helped us classify the images as normal and abnormal automatically extracting features from the whole image presented. The two CNN models implemented with RPN have VGG16 and ResNet50. The Region Proposal Network (RPN) with CNN (as explained in Section 2) has an improved abnormality identification ability due to ROI localization and feature extraction at an intermediate layer. RPN with VGG16 for object identification and classification recorded an average accuracy of 83.1%. When ResNet50 was used the metric achieved 85.73%. It is graphically represented in Fig. 3 below.

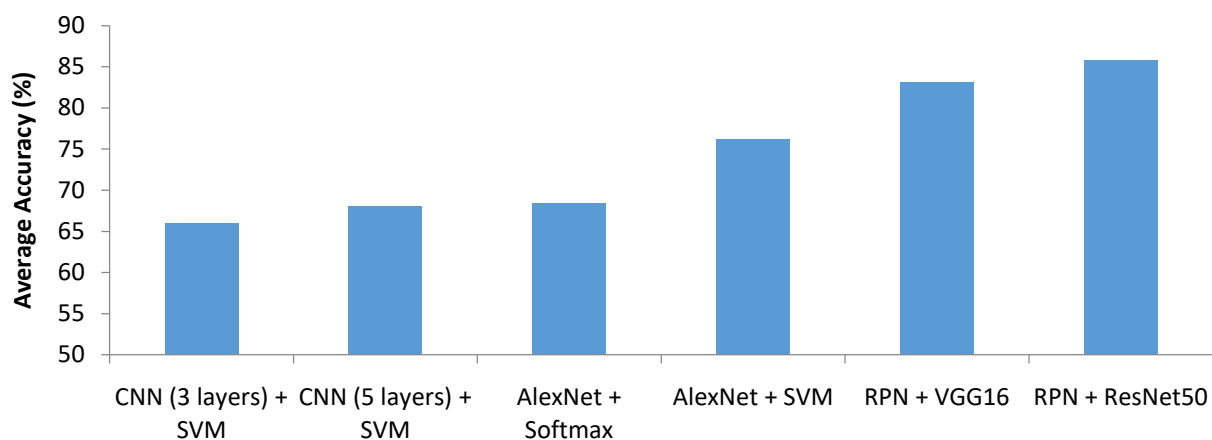


Fig. 3 Training Accuracies for the various models

For classification, sparse cross-entropy was used and Smooth L1 lossfunction was used for bounding box. Total losses using the Equation 3 was obtained for both the networks in different epochs. Region Localization or L_{BOX} loss for Region Proposal Model was shown in Fig.4 for VGG16 and ResNet50 respectively. The X- axis gives the number of epochs and the Y-axis represents the loss.

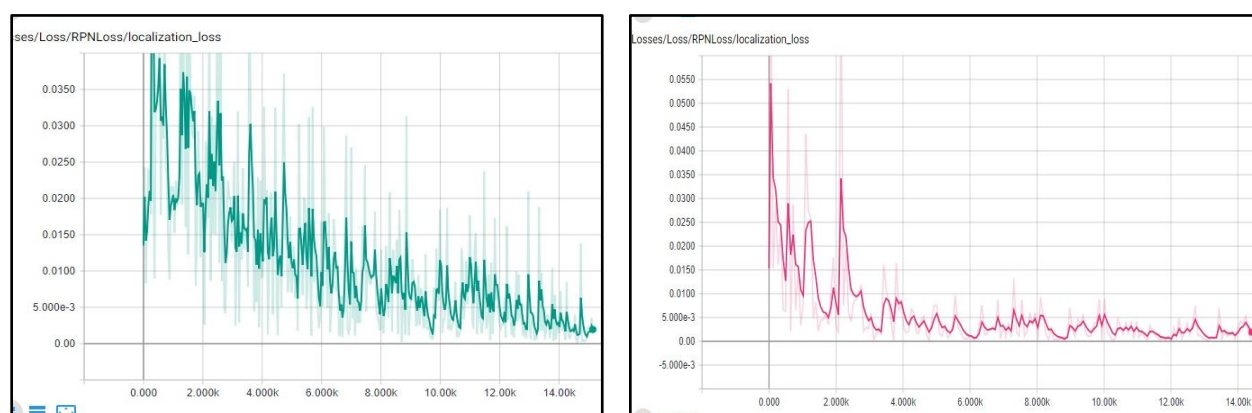


Fig.4: Localization Loss for Bonding Box from Region Proposal Model

It can be observed from the graphs above that ResNet50 converged much faster in 3K epochs and more accurate bounding box were returned. Similarly, bounding box proposals at classification stage result in another set of Regression Loss L_{BOX} . Fig.5 give the respective Bounding Box classification loss for the above networks. The X- axis gives the number of epochs and the Y-axis represents the loss. It can be seen that the losses converge faster initially and final convergence was achieved in 3K epochs in ResNet50 as compared to VGG16. We can now say that for L_{BOX} losses in both detector and classifier network converge faster in ResNet50.

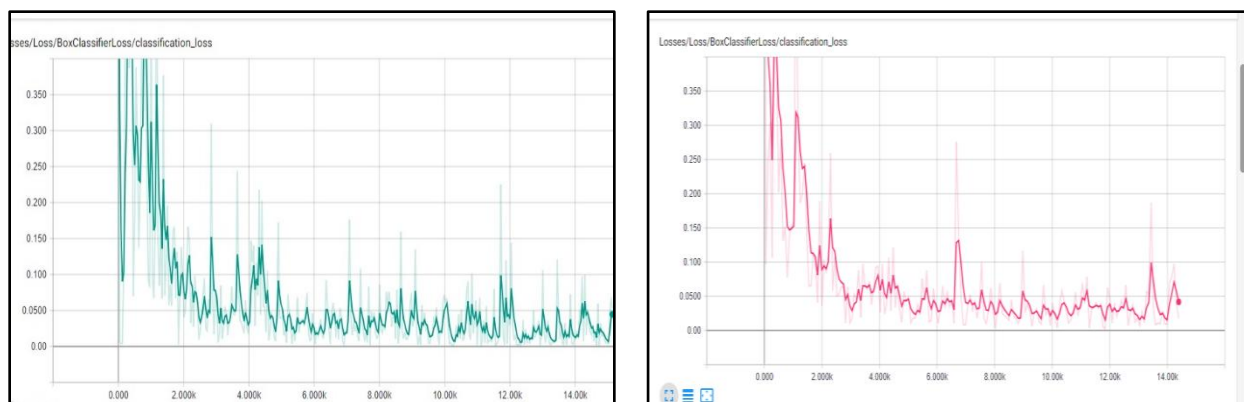


Fig. 5: Bounding Box Regression Loss from Detector Model

Classification losses or L_{CLS} (computed using Equation 4) at the object detector network result in classification of a region for presence of an object (foreground) or the absence (background). It is shown below in Fig. 6 for the two networks respectively. The X- axis gives the number of epochs and the Y-axis represents the loss.

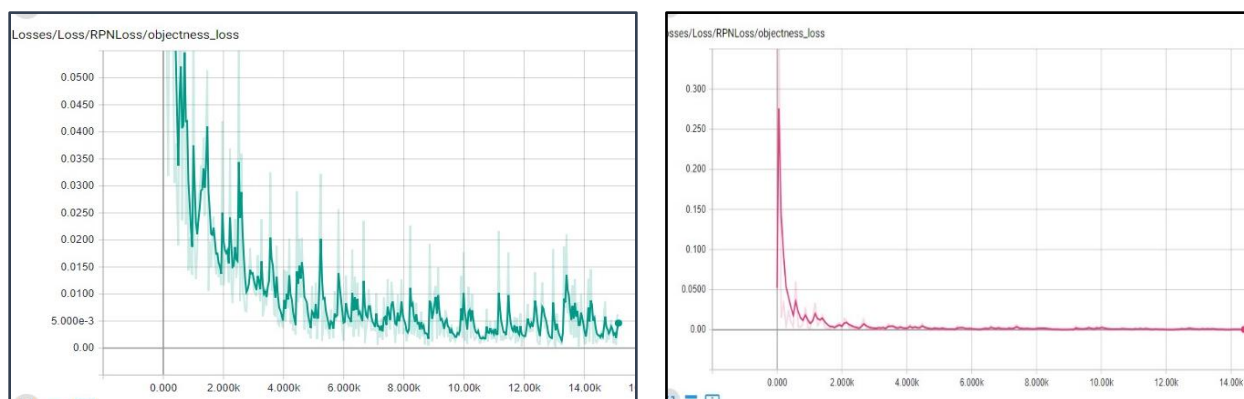


Fig.6: Objectness Loss or Region Classification Loss from Region Proposal Model

The region classification was accurately done in 2K epochs only with little changes later. Fig. 7 gives the L_{CLS} comparison of the two networks at the classification stage. The X- axis gives the number of epochs and the Y-axis represents the loss.

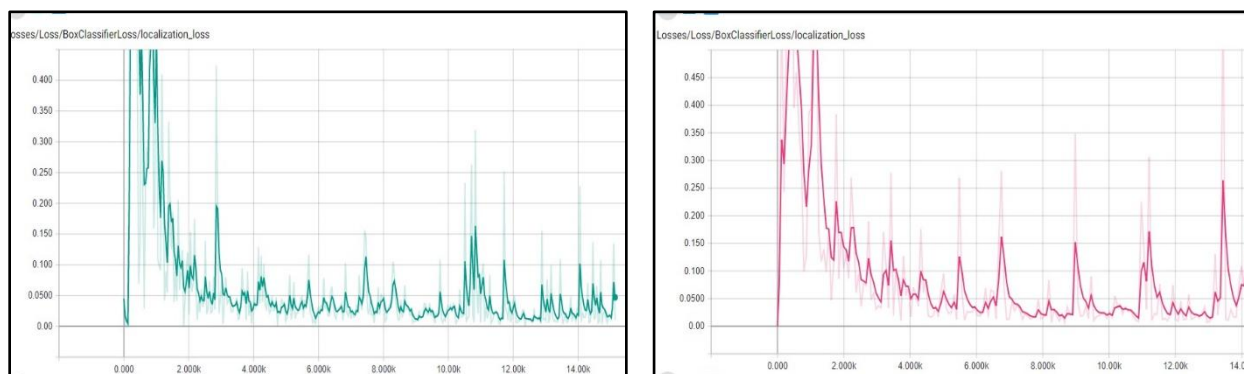


Fig.7 Classification Loss from Detector model

Classification loss in ResNet50 converged after 5K epochs. A high probability of presence of the object in the region was signified. Summing up all the losses giving more weightage to

classification losses as compared to regression losses gives us the Total Loss function of the models as shown in Fig.8. The X- axis gives the number of epochs and the Y-axis represents the loss.

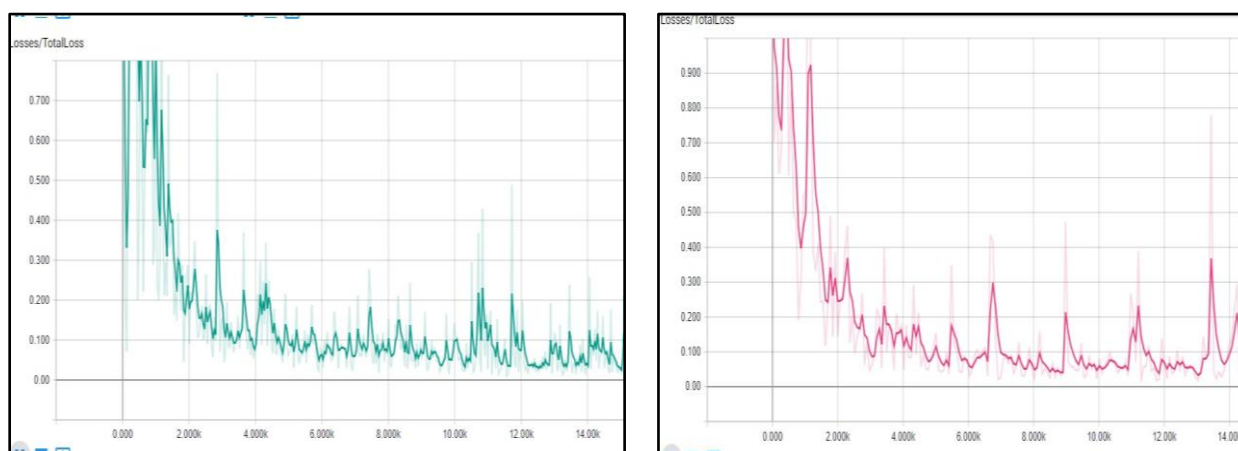


Fig.8: Total Loss from classification and regression of Proposal and Detector Networks

Total Loss in RPN+ResNet50 started converging at 5K epochs, much earlier than RPN+VGG16 model and this can be verified in Fig. 8 above. ResNet50 performs better for region localization and objectness loss. Our accurately generated bounding box can actually mark the Region of Interest and assist the ultrasonologists /clinicians. When implemented with VGG16, the average training accuracy was 83.1% whereas with ResNet50, average accuracy was 85.73%.

4.3 Predictive Performance Analysis

Trained models were experimented with test TVS images of 17 new patients to perform the predictive analysis. The predictive performance of the different models were captured from which the True Positives, True Negatives, False Positives, False Negatives were recorded. Different metrics used to analyze the performances are as shown in the Table 1. Sensitivity describes proportion of patients having TB being correctly identified as having TB whereas Specificity gives the proportion of non - TB cases that were correctly predicted. Sensitivity is more important here as we do not want to miss a TB patient in the preliminary examination whereas if a non-TB case is classified wrongly, the correlating symptoms and further investigations would rule out such cases. F1 measure is a single measure that considers sensitivity and percentage of correctly predicted TB cases.

CNN (3 layers) and CNN (5 layers) with SVM classifier when compared show that there was an overall improvement in all the metrics when the number of layers was increased. This means deeper the network, better the feature extraction as more complex features are extracted. AlexNet with 8 layers, further increased the accuracy but sensitivity decreased. This was not desirable, and on changing the classification from Softmax to SVM there was an increase in all metrics.

Table 1: Performance Measures of various Deep Learning Architecture on the dataset

Architectures	Mean Accuracy(Testing)	Sensitivity	Specificity	F1-measure
CNN (3 layers) + SVM classifier	0.607	0.67	0.55	0.61
CNN (5 layers) + SVM classifier	0.666	0.72	0.62	0.66

AlexNet + Softmax Layer	0.677	0.63	0.71	0.66
AlexNet + SVM classifier	0.74	0.77	0.71	0.74
RPN+ VGGNet-16	0.818	0.81	0.824	0.81
RPN+ ResNet50	0.849	0.88	0.813	0.851

Further, RPN with deeper CNN's, VGG16 and ResNet50 showed better accuracy, sensitivity, specificity and F1 measure. Though the accuracy of ResNet50 turned out to be better but the specificity of VGG16 was better. With the analysis above, RPN with ResNet50 executed for 4K epochs only was able to identify abnormal image with maximum accuracy as compared to other models. ResNet50, though deeper is less complex and trains faster than VGG16. The mean accuracy also increased by 3.1% when ResNet50 was used. A F1 score of 0.851 and the sensitivity measure of 88% was observed.

Sample output with identified bounding box around the object and classification as normal or abnormal with a confidence score of classification by ResNet50 are shown in Fig. 9.



Fig. 9: Identified Endometrial TB region with a Bounding Box on Test Images using RPN+ResNet50.

4. CONCLUSION

Genital TB is one of the major contributors to cause of rising female infertility in India and other developing countries, but reported cases are much less as the symptoms are varied and at times non-existent making the diagnosis pretty challenging. In this paper, a computational method using deep learning is proposed for abnormality localization and identification of endometrial tuberculosis in the presented ultrasound images. In the network model, RPN has been used to localize the tubercular region and CNN with ResNet50 is used for prediction. The model has yielded a mean predictive accuracy of 84.9% and a F1 score of 0.851 for the dataset in hand. Further, the sensitivity shown by the model is 88%.

Deep Learning for ultrasound imaging to identify Endometrial TB in infertility patients could prove to be a major breakthrough. If TB is the reason of infertility, as it is in most of the cases, then this early detection during the initial phases of investigation will be a great relief. In this paper, we have used only ultrasound image analysis using a deep neural network model; in the future, we are planning to integrate medical expertise and knowledge to further improve the model efficiency.

FUNDING

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

REFERENCES

1. WHO,WHO global tuberculosis report 2016.Available from:http://www.who.int/tb/publications/global_report/en/
2. Akkus, Z., Cai, J., Boonrod, A., Zeinoddini, A., Weston, A. D., Philbrick, K. A., & Erickson, B. J. (2019). A Survey of Deep-Learning Applications in Ultrasound: Artificial Intelligence–Powered Ultrasound for Improving Clinical Workflow. *Journal of the American College of Radiology*, 16(9), 1318-1328.
3. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., andAsari, V. K. 2018. The history began from alexnet: A comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164.
4. Aslan, G., Ulger, M., Ulger, S. T., Durukan, H., Yazici, F. G., andEmekdas, G. 2018. Female genital tuberculosis cases with distinct clinical symptoms: Four case reports. *International Journal of Reproductive Biomedicine*, 16(1). 57.
5. Byrne, A. L., Marais, B. J., Mitnick, C. D., Lecca, L., Marks, G. B, 2015. Tuberculosis and chronic respiratory disease: a systematic review. *International Journal of Infectious Diseases*.32.138-146.
6. Cahan, A., Cimino, J. J, 2017. A learning health care system using computer-aided diagnosis. In: *Journal of medical Internet research*. 19(3), pp. e54.
7. Chen, H., Ni, D., Qin, J., Li, S., Yang, X., Wang, T., &Heng, P. A. 2015. Standard plane localization in fetal ultrasound via domain transferred deep neural networks. *IEEE journal of biomedical and health informatics*, 19(5), 1627-1636.
8. Chi, J., Walia, E., Babyn, P., Wang, J., Groot, G., &Eramian, M. 2017. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. *Journal of digital imaging*, 30(4), 477-486.
9. De Bruijne, M., 2016. Machine learning approaches: From detection to diagnosis. *Medical Image Analysis*. 33. 94-97.
10. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., &Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: *IEEE conference on computer vision and pattern recognition*. pp. 248-255.
11. Doi, K., 2007. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. In: *Computerized medical imaging and graphics*. 31(4-5). 198-211.
12. Grace, G. A., Devaleenal, D. B., andNatrajan, M., 2017.Genital tuberculosis in females.*The Indian journal of medical research*, 145(4). 425.
13. Grossman, M. (2019). Proposal networks in object detection, Dissertation.
14. Han, S., Kang, H. K., Jeong, J. Y., Park, M. H., Kim, W., Bang, W. C., &Seong, Y. K. 2017. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Physics in Medicine & Biology*, 62(19), 7714.
15. He, K., Zhang, X., Ren, S., & Sun, J. 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
16. He, K., Zhang, X., Ren, S., & Sun, J. 2016.Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*.770-778.
17. Huang, Q., Zhang, F., & Li, X., 2018. Machine learning in ultrasound computer-aided diagnostic systems: a survey. In: *BioMedicalResearch International*.1-11.

18. Huang, Y. L., Chen, D. R., Jiang, Y. R., Kuo, S. J., Wu, H. K., & Moon, W. K. , 2008. Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound. *Ultrasound in Obstetrics and Gynecology*.*The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 32(4).565-572.
19. Jacobson, J. L., 2018. Women's health: The price of poverty. In: *The Health of Women*. 3-32.
20. Jindal, U. N., 2006. An algorithmic approach to female genital tuberculosis causing infertility. *The International Journal of Tuberculosis and Lung Disease*, 10(9).1045-1050
21. Jirge, P. R.2016.Poor ovarian reserve. *Journal of human reproductive sciences*, 9(2), 63.
22. Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C., 2019. A Novel Deep Learning based Framework for the Detection and Classification of Breast Cancer Using Transfer Learning. In: *Pattern Recognition Letters*. 125.1-6.
23. Khan S. and Shamsi J.A, 2019. Health Quest: A generalized clinical decision support system with multi-label classification. In: *Journal of King Saud University – Computer and Information Sciences*.
24. Khurana, A.,Sahi, G., 2013. OC14. 04: Ultrasound in female genital tuberculosis: A retrospective series. In: *Ultrasound in Obstetrics & Gynecology*, 42(s1).28.
25. Krizhevsky, A., Sutskever, I., & Hinton, G. E., 2012.ImageNet classification with deep convolutional neural networks.In *Advances in neural information processing systems*.1097-1105.
26. Legro, R. S.,Hurtado, R. M., Kilcoyne, A., Roberts, D. J., 2016. Case 28-2016: A 31-Year-Old Woman with Infertility. *New England Journal of Medicine*, 375(11).1069-1077.
27. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Li, S. X.& Wang, T.2019. Deep learning in medical ultrasound analysis: A review. *Engineering: Research AI for Precision Medicine—Review*. 5(2). 261-275
28. Mansour, R. F., 2018. Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy. *Biomedical engineering letters*, 8(1). 41-57.
29. Meng, D., Zhang, L., Cao, G., Cao, W., Zhang, G., & Hu, B. (2017). Liver fibrosis classification based on transfer learning and FCNet for ultrasound images. *IEEE Access*, (5), 5804-5810.
30. NamavarJahromi, B., Parsanezhad, M. E., and Ghane-Shirazi, R., 2001. Female genital tuberculosis and infertility. In: *International Journal of Gynecology & Obstetrics*, 75(3).269-272.
31. Nugroho H.A, Zulfanahari, Frannita E.L, Ardiyanto I., Choridah L.,2019. Computer Aided Diagnosis for thyroid cancer system based on internal and external characteristics. *Journal of King Saud University-Computer and Information Sciences*,doi.org/10.1016/j.jksuci.2019.01.007
32. Pan, S. J., & Yang, Q., 2009.A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*. 22(10).1345-1359.
33. Patil, A. V., Somasundaram, K. V., andGoyal, R. C, 2002.Current health scenario in rural India. *Australian Journal of Rural Health*, 10(2).129-135
34. Razavi, F., Tarokh, M. J., &Alborzi, M., 2019.An intelligent Alzheimer's disease diagnosis method using unsupervised feature learning. In: *Journal of Big Data*. 6(1).32.

35. Ren, S., He, K., Girshick, R., & Sun, J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*.91-99.
36. Ren, S., He, K., Girshick, R., Zhang, X., & Sun, J.2016. Object detection networks on convolutional feature maps. *IEEE transactions on pattern analysis and machine intelligence*, 39(7), 1476-1481.
37. Sanches, I., Carvalho, A., & Duarte, R., 2015. Who are the patients with extrapulmonary tuberculosis? *Revista Portuguesa de Pneumologia (English Edition)*, 21(2).90-93.
38. Sharma, J. B., 2015. Current diagnosis and management of female genital tuberculosis. *The Journal of Obstetrics and Gynecology of India*, 65(6),362-371.
39. Sharma JB, Dharmendra S, Agarwal S, Sharma E. 2016 Genital tuberculosis and infertility. *FertilityScienceandResearch*3(1),6-18
40. Sharma, J. B., Sharma, E., Sharma, S., &Dharmendra, S., 2018. Female genital tuberculosis: Revisited. *The Indian journal of medical research*, 148(Suppl 1), S71.
41. Simonyan K. and Zisserman A., 2015. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations*.arVix:1409.1556v6.
42. Sheoran, P. and Sarin, J., 2015. Infertility in India: social, religion and cultural influence. *International Journal of Reproduction, Contraception, Obstetrics and Gynecology*, 4(6), 1784.
43. Shung, K. K., 2011. Diagnostic ultrasound: Past, present, and future. *Journal of Medical and Biological Engineering*,31(6).371-4.
44. Song, H., Nguyen, A. D., Gong, M., & Lee, S., 2016. A review of computer vision methods for purpose on computer-aided diagnosis. *Journal of International Society for Simulation Surgery*. (3).1-8
45. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D&Rabinovich, A. 2015.Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*.1-9.
46. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C, 2018.A survey on deep transfer learning. *International Conference on Artificial Neural Networks*.270-279.
47. Yassin, N. I., Omran, S., El Houbby, E. M., &Allam, H.,2018. Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer methods and programs in biomedicine*,156, 25-45.