Proceedings of 13th IOE Graduate Conference
Peer Reviewed
ISSN: 2350-8914 (Online), 2350-8906 (Print)
Year: 2023 Month: April Volume: 13

Diagnosis of Malignant Thyroid Nodule of Ultrasound Image using VGG-19

Madhu Kumari Jha [a], Hom Nath Tiwari [b]

[a] *Department of Electronics & Computer Engineering, Paschimanchal Campus, IOE, Tribhuvan University, Nepal*
[a] madhujha2050@gmail.com, [b] homnath@wrc.edu.np

Abstract
Globally detection of thyroid disease has become to be a serious medical problem in which thyroid nodules are relatively common and rarely cancerous which requires efficient automatic prediction models. The available techniques such as ultrasonography imaging and percutaneous biopsy focuses on binary classification optimization model with small size dataset without validating results whether a nodule is benign or malignant, and still required well experienced and senior radiologist. This paper focuses to develop customize model network using Visual Geometry Group-19 (VGG) architecture system which also determines the percentage of the infected area with decreasing the unwanted redundant data from the patient's database. 480 images in the grayscale are used from open access Thyroid Digital Image Database (TDID) . The customized VGG 19 architecture has classified and diagnosed 97% of positive cases correctly with 95% accuracy which is efficient than the manual diagnosis. The analysis shows that the level of accuracy is limited by volume of dataset.

Keywords
Thyroid detection, Malignant nodule, VGG 19, suspicious nodule, Deep learning

1. Introduction

Thyroid nodule is a kind of lump which get developed in thyroid gland. It can be in solid form or filled with fluid in single or in cluster form. Relatively thyroid nodules are common in general and rarely cancerous. Recent study shows that Thyroid disease is more common among people aged 16 to 30 and then 41 to 45 across all age groups [1]. Thyroid diseases, including hyperthyroidism Graves' disease, Hashimoto's thy- roiditis or hypothyroidism with autoimmune and postpartum thyroiditis are among the pathophysiology that link to autoimmune [2]. Thyroid nodule is in two categories such as malignant nodules that can cause thyroid cancer and benign nodules that are non-cancerous which is based on their characteristics. In both situations, thyroid nodules can negatively impact a patient's health. The physical examinations, thyroid function testing, and fine needle aspiration (FNA) biopsies are the ways to diagnose a thyroid nodule [3]. These methods does not give deep information about the nodule's condition instead gives a primary test evaluation. These laborious procedures are prone to mistakes, which put the patient's life in danger. Hence a quicker and more accurate method is needed to identify and categorize thyroid nodules.

This paper implements the thyroid ultrasonography image classification such as benign nodules as Thyroid Imaging Reporting & Data System 2 (TIRADS 2) and TIRADS 3 and malignant nodules as TIRADS 5 and TIRADS 4a, 4b, and 4c. Figure 1 and Figure 2 illustrate TIRADS 2 and TIRADS 4a respectively. It is done manually by the physician which is tedious and might get wrongly diagnosed. So in this paper, method is automated by using and assessing a malignant thyroid classification technique. Fundamentally, the categorization is done using Convolutional Neural Network (CNN), which is based on the VGG-19 architecture.

The details of this paper are organized as: Part 1: gives the background and introduction of this paper work, Part 2 explains related works done in the areas of thyroid and VGG-19 as well as the research gap in those initiatives. The techniques used are described in Section 3. The results of the models' performance are displayed in section 4. The conclusion is provided in section 5.

2. Related Works

In current scenario the thyroid disorders have huge effect in the population worldwide. The alternative development is necessary for improving the diagnosis process. Thyroid cancer detection and thyroid disease detection has been done by machine learning or deep learning methods in research so far.

The study shows, the VGG-19 model, which is pre-trained on the ImageNet dataset, outperformed ResNet50, Inceptionv3, and EfficientNetB0, in terms of classification performance. The ensemble method performed well for all thyroidal image diagnosis, classes in terms of recall and precision. An efficient, lightweight CNN model and a pre-trained, finely tailored VGG-19 model are combined to create a CNN- VGG ensemble [2]. Similar to this, a deep learning-based thyroid-breast nodule CAD ultrasound diagnostic system is proposed in CNN for Breast and Thyroid Nodules Diagnosis in Ultrasound Imaging . Which shows Positive predictive value (95.41%), negative predictive value (98.05%), and total AUC (0.96) indicate good performance [3]. The Random Forest (RF) shows the best performance accuracy of 94.8% for classifying thyroid disease among Decision Trees (DT), RF, K-Nearest Neighbors classifier (KNN), and Artificial Neural Network (ANN) [4]. Whereas Deep Neural Network (DNN) methods predict thyroid illness categorization and achieves accuracy of 99.5% but requires a large dataset to train the model with sampled and unsampled

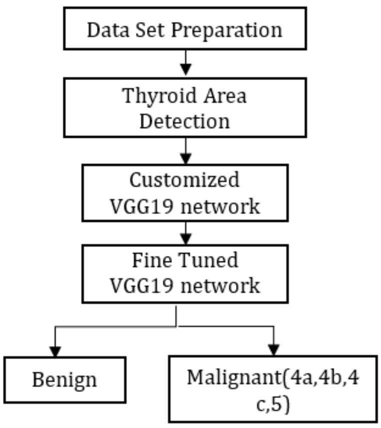

Figure 1: Block Diagram of applied Methodology

data which again requires additional computational resources as a drawback [5].

3. Methodology

Clean dataset is initial requirement to be used in machine learning algorithm for which dataset preparation becomes the essential part in the process. The deep learning-based system is created by merging the categorization outcomes of various CNN networks. Figure 1 illustrate the proposed networks followed by series of steps and process.

Here, first fine-tuned the CNN architectures that has been pre-trained (initialized) on natural image data. The purpose is to classify the thyroid tumors on ultrasound images based on six different categories:

- 1(Benign)
- 2(Benign)
- 4a (Malign)
- 4b (Malign)
- 4c (Malign)
- 5 (Malign)

The system categorizes if the input image is cancerous or non-cancerous that is benign or malign. In general, the dataset available is imbalanced, and several images contain thyroid cancers that are not labeled.

Also, some images have two thyroid ultrasound images from the same subject. Before algorithm training, images are processed. In data preparation, firstly images are normalized, secondly images are cropped to find the biggest contour of image. Whereas some images contain text about classification of thyroid tumors and required Pre- processing in refining image data at initial stage so that it may be utilized to process data more effectively. Augmentation is used which tries to avoid overfitting even at micro variations which consists of a random horizontal flip, a random rotational distance of 20, a random zooming area in the range of 0.2, and a random contrast in the range of 0.1. In this work the convolutional(conv.) layer operation is explained as (1).

$$FM(i,j) = (I * F)(i,j) = \sum\sum I(i+m, j+n)F(m,n) \quad (1)$$

I is the input matrix, F denotes a 2D filter of size (m, n), I * F stands for the convolutional operation, and FM denotes the output of a 2D feature map.

3.1 Implemented VGG-19 network

VGG-19 architecture which is pre-trained Neural network have been applied in this work Figure 2. At first, the last layer of the model has been removed and made non-trainable all the layer of model. Freeze the original layers so the model's weights could not be further optimized and the model's previous layers' acquired lessons could not be adjusted to the new data set. Now Flatten layer have been added on model output for obtaining a 1-D array of features.Fully Connected(FC) layer with 256 hidden units is added and one dropout is performed. The FC layer reduces the number of neurons in each step of the layer, which increases the independence of the neurons from their neighbors and preventing overfitting. The likelihood of a neuron activating is measured by the rate parameter, which is set to 0 and dropping out the neuron for inactive neurons in the previous layer. The last addition is completely integrated layer with four hidden units and a SoftMax activation algorithm. The model is then trained for several epochs, optimized using the Adam optimizer, and the final model is obtained.

3.2 Fine tuning of VGG-19 modal

Figure 2 output is being input for the finetuned segment. Freezing the convolutional base and feeding the output to the diagnosis classifiers was a fine-tuning strategy. Deep layer models that have already been trained on a certain dataset are subjected to fine tuning. The final set of fully linked layers from the pre-trained network are removed using this technique, and they are replaced with a fresh set of fully connected layers with random initializations. Block5-pool is transfer layer whose output is redirected to a new fully connected neural network that does the categorization. The original pre-trained model is frozen during training of the new classifier in transfer learning. This guarantees that the original VGG-19 model's weights are not changed. New classifier causes its training, not to introduce significant gradients into the VGG-19 model, which could alter its weights or lead to overfitting to the new dataset. Once the new classifier is trained, the deeper layers of the VGG-19 model as well as the networks are also tweaked. The original VGG-19 model's convolutional layer and all pertinent layers are then enabled for the trainable Boolean. The final two convolutional layers, whose names include "block4" and "block5" are trained. Now, it is necessary to

unfreeze the layer and recompile it using the Adam optimizer along with the new classifier to fine the adjustment. The input to VGG-19 is downsized (224 px, 224 px) scale images during training. A stack of convolutional filters with a very narrow receptive field 3X3, are utilized to process the input image. For 3X3 convolution layers, the padding is the same. With stride 2, max-pooling is carried out over a 2X2 pixel window. Three FC layers: the first two of which contain 1024 channels each and the soft- max layer in between, come after a stack of convolutional layers. In all hidden layers, the rectification (ReLU) non-linearity is used. L1-L2 regulazier has been utilized to prevent overfitting in FC layers. The weighted binary cross-entropy loss function is used. Also, the model is trained with 40,50,100,200,500 epochs and 8,16,32,128 batch sizes.

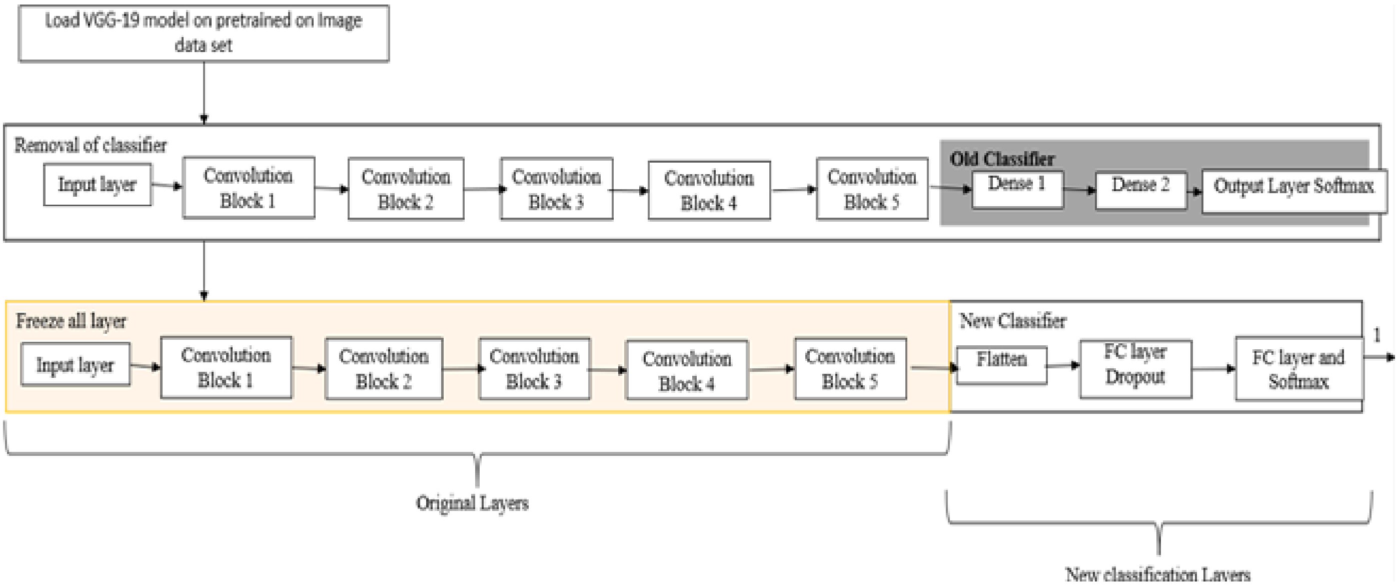

Figure 2: Applied VGG-19 Based Classification Block Diagram

3.3 Dataset

The dataset has been collected from open access database for thyroid nodule TDID, which contains in total 480 valid cases and the images is in the grayscale. Among the 480 cases with TIRADS score, 280 cases were diagnosed as malignant (TIRADS score 4a, 4b, 4c and 5) and 200 cases as Benign (TIRADS score 1and 2). The image augmentation process is used to produce 2000 number of datasets for training the CNN model. Among them, 1400 images are used as training, 400 image for validation and rest 200 images for test sets.

4. Result and Discussions

4.1 Identification of Suspicious Nodular Areas

Two bounds horizontal projection and vertical projection are employed to identify suspicious thyroid regions and detect suspicious nodules, which eliminates unnecessary image artifacts. Figure 4 shows that the image is first separated into three sections: the skin, the thyroid area, and the dark region to extract the anatomical information from it. Thyroid area intensity is between skin and dark region with skin having higher gray levels. The thyroid area is divided from other areas using a horizontal projection in accordance with the Otsu's threshold criteria. In Each row of the images, the average intensity is determined by using Otsu's Thresholding and separate the intensity information for both skin parts and nodules parts. Finally, the threshold omits the high intensity part and the average intensity part shown by a dark spot which indicates the nodular area in the given ultrasonography image. The dark spot shows the thyroid affected area in Figure 5.

4.2 Parameter Tuning

4.2.1 Effects of learning rates on Accuracy:

The accuracy and loss value at every training step have been recorded in a log file and then all values are plotted. Figure 6 shows that, training with 0.1 and 0.01 learning rate does not converge which results low training accuracy and high validation losses. Whereas training with learning rate 0.001 results high training accuracy and low validation losses.

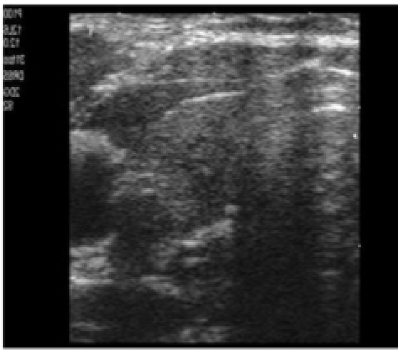

Figure 3: Original Image of Thyroid Nodule(TIRADS5)

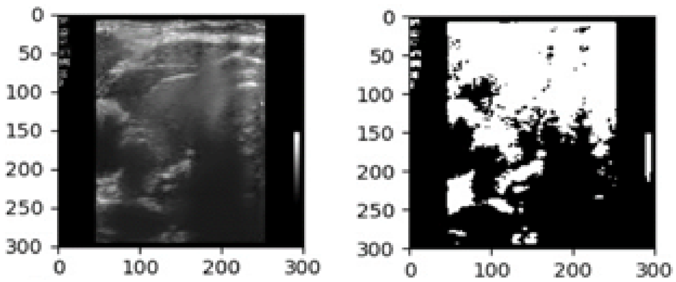

Figure 4: Suspicious Nodule area Detection using Image Thresholding

4.2.2 Effects of Batch Sizes on Accuracy:

The VGG-19 network is trained through the ultrasonography Image of training dataset, varying by batch size as 32, 64 and 128. The accuracy and loss value at every training step is recorded in a log file and then all values are plotted in Figure 7. That shows, training with 32 and 64 batch size converged and result high training and validation accuracy whereas batch size with 128 results high training and validation loss and does not converge.

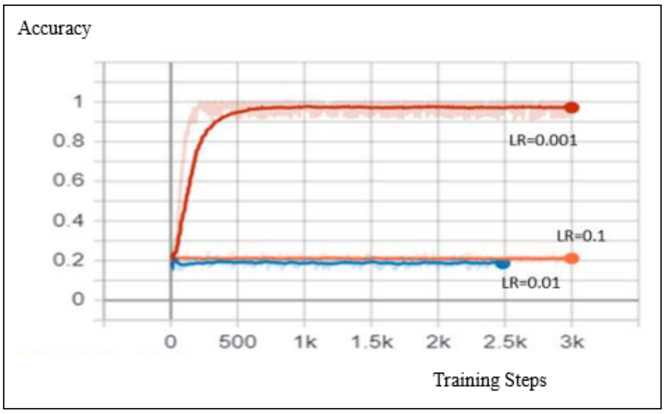

Figure 5: Accuracy at various learning rate

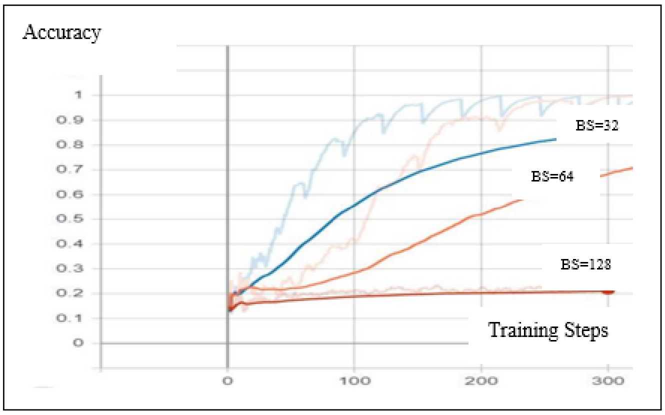

Figure 6: Accuracy at various Batch sizes

4.3 Test Result

Among the test samples from the customized VGG-19 network, the entire network is trained with 1200 training data for 100 epochs. Here two results are compared, one is by bare eyes, that's by doctors and another is by proposed system. Figure 9 proposed system classified as Malignant with 92.28% of Malignant nodules. Originally same US image has been diagnosed as TIRADS 4c category which also explain that the person has cancerous nodule in his thyroid gland.

Figure 10 originally diagnosed as TIRADS 2 by the specialist doctors and the same image is classified Benign with 99.23%

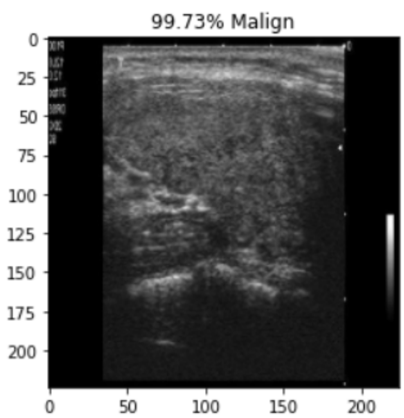

Figure 7: Classified as Malign with 99.73% of malignant nodule

that is non-cancerous nodules by proposed system Figure 11. Similarly, the proposed system Classified the malign with 90.23% malignant nodules and non-cancerous with 68.35% benign nodules in Figure 12, Figure 13 respectively. The difference in classification done by senior radiologist of TDID dataset and the model Table 1. Among 200 cases, 60 cases are classified as the benign and 120 cases are classified as the malignant nodule Table 1.

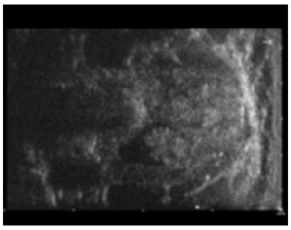

Figure 8: Originally diagnosed as TIRADS 4c

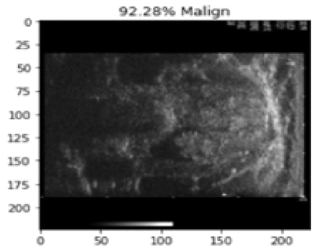

Figure 9: Classified as Malign with 92.28% of malignant nodules

Figure 10: Originally diagnosed as TIRADS 2 by doctors

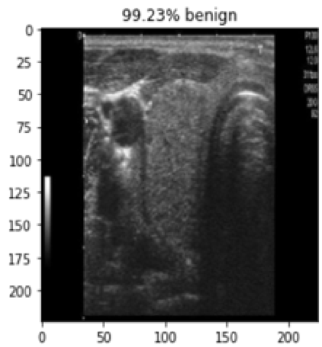

Figure 11: Classified as benign with 99.23% of benign nodules

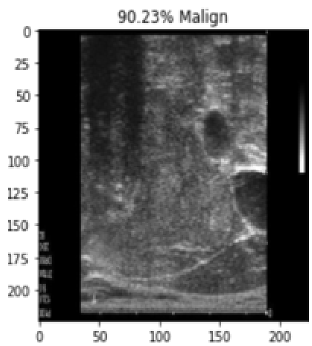

Figure 12: Classified as Malign with 90.23% of Benign nodules

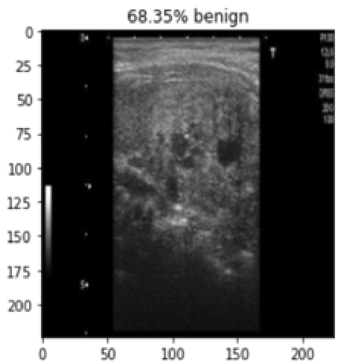

Figure 13: Classified non-cancerous with 68.35% of Benign nodules

Table 1: Classification Performance of the given Fine-tuned VGG-Network

Thyroid Type	Classification by Radiologist (Based on TDID dataset)	Classification by VGG-19
TIRADS 2	33	Benign (60)
TIRADS 3	32	
TIRADS 4a	43	Malignant (120)
TIRADS 4b	28	
TIRADS 4c	33	
TIRADS 5	32	

4.4 Receiver operating characteristics (ROC) and Confusion Matrix

The ROC curves of model computed for each diagnosis class showed the rate of false-positive is in between 0.1 to 0.2 while the rate of true positive is between 0.9 to 1 as in Figure 14.The experiment showed that the proposed model can be good and stable classifier for malignant thyroid images. The confusion matrix explains here the correctness and accuracy of model. Confusion matrix represent that the system could not classify the thyroid ultrasonography image hundred percent but the overall performance score of the model is satisfactory at 95% in Figure 15.The system correctly categorizes 160 malignant nodules thyroid images and 20 wrongly classified malignant thyroid images. Similarly, the system correctly classifies 180 images to benign and 17 wrong images to benign.

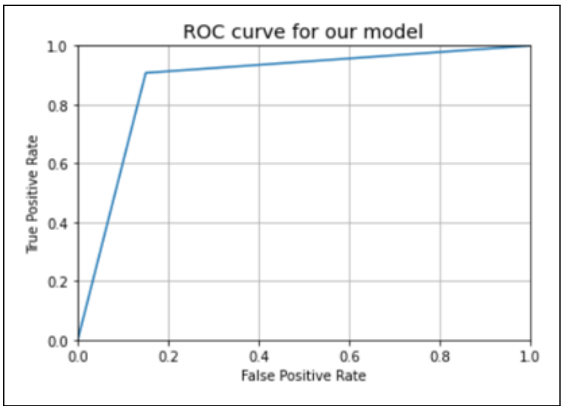

Figure 14: ROC curve

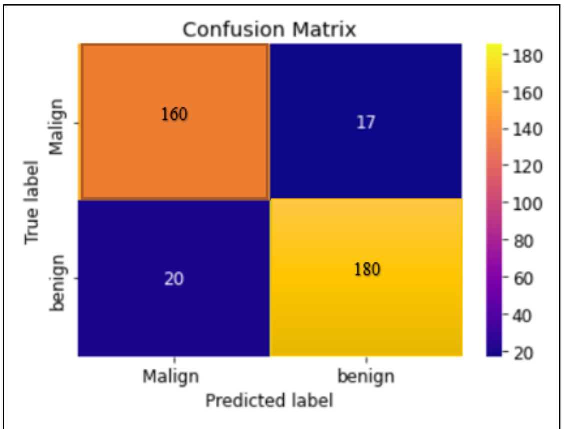

Figure 15: Confusion Matrix

4.4.1 Accuracy and Loss plot

The recorded history during training network of accuracy on each epoch is used to get plot of accuracy metrics. For visualizing, the accuracy of the training data ("acc") and the validation data ("val-acc") has been chosen. The accuracy of model is plotted for epoch 100 and the system Loss is plotted for 100 epochs in Figure 16 and Figure 17 respectively.

4.4.2 Classification Reports

Table 2 shows the data seen as the average range of classifications of malignant thyroid nodule during the process. In proposed architecture, the recall level is 0.97 for 200 epochs, 0.93 for the 100 epochs 0.97 for 50 and 0.91 for 40 with highest accuracy 0.95 for epochs 100. 200 epochs are the point of converging, as above 200 epochs, over-fitting case gets occurred.

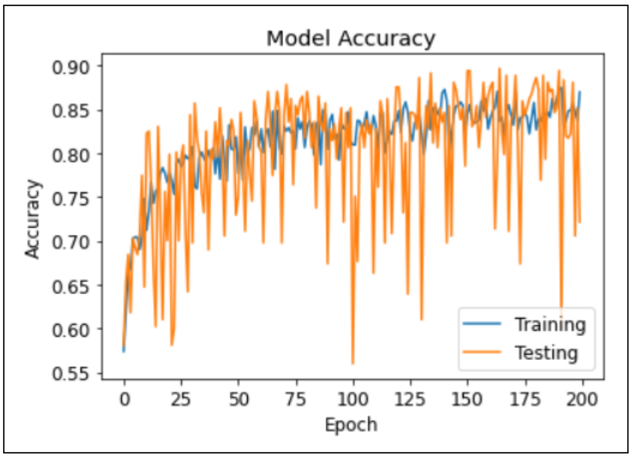

Figure 16: Model Accuracy for 200 epochs

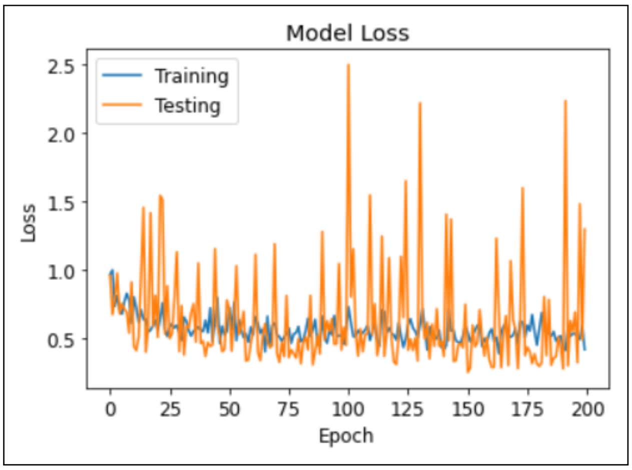

Figure 17: Model Loss for 200 epochs

Table 2: Classification reports

Epoch	Recall	F1-score	Precision	Accuracy
200	0.97	0.85	0.95	0.86
100	0.93	0.94	0.93	0.95
50	0.97	0.84	0.94	0.91
40	0.91	0.88	0.91	0.88

5. Conclusion

Thyroid detection has become to be a serious medical problem which needs efficient automatic prediction models. TIRADS and US-guided FNA have been demonstrated to significantly increase the accuracy of identifying malignant thyroid nodules, whereas conventional biopsy is exceedingly time- consuming and harmful to the human body. In this work, new customized network is developed using the existing VGG-19 architecture for ultrasound thyroid image. The Suspicious area in image is classified as either it is malignant or not which shows that the more the percentage of infected area, the more the serious case of cancer. Result shows that TIRADS 2 and TIRADS 3 are classed as benign, but TIRADS 4a, 4b, 4c, and TIRADS 5 are shown to have an image of malignant nodules. The developed system works better for malignant diagnosing by classifying the infected area. For 200 epochs the result obtained on test dataset overall diagnosis Recall/sensitivity of 97%, precision of 95% and accuracy of 86% similarly for 100 epoch the result obtained is 93% Recall, 93% precision and 95% accuracy which is more efficient than manual diagnosis and more than sensitivity/recall of 90.844%, precision of 94.865% and accuracy 96.1% in single VGG-19 model and recall of 95.752%, precision of 95.41% and accuracy of 97.352% in CNN-VGG ensemble model [2].

Further, this system also decreases the unwanted redundant data from the patient's database and makes cost effective and have high speed of performance. The proposed system successfully diagnoses the percentage of infected area and mainly focuses on identifying the malignant thyroid nodules by making the treatment of thyroid cancer patient simple by reducing further complex procedure in convenient way. The analysis shows that the level of accuracy is limited by volume of dataset.

References

[1] Anup Shamsher Budhathoki, Suprita Gupta, Sanjay Kumar Sah, Navin Kumar Sah, Nirdhan Yadav, Amit Kumar Patel, Afroj Ansari, and Alan Pandey. Status of thyroid function among patients of national medical college and teaching hospital. *Journal of National Medical College*, 5(1):64–70, 2020.

[2] Corina Maria Vasile, Anca Loredana Udriștoiu, Alice Elena Ghenea, Mihaela Popescu, Cristian Gheonea, Carmen Elena Niculescu, Anca Marilena Ungureanu, Ștefan Udriștoiu, Andrei Ioan Drocaş, Lucian Gheorghe Gruionu, et al. Intelligent diagnosis of thyroid ultrasound imaging using an ensemble of deep learning methods. *Medicina*, 57(4):395, 2021.

[3] Xiaowen Liang, Jinsui Yu, Jianyi Liao, Zhiyi Chen, et al. Convolutional neural network for breast and thyroid nodules diagnosis in ultrasound imaging. *BioMed Research International*, 2020, 2020.

[4] Tahir Alyas, Muhammad Hamid, Khalid Alissa, Tauqeer Faiz, Nadia Tabassum, and Aqeel Ahmad. Empirical method for thyroid disease classification using a machine learning approach. *BioMed Research International*, 2022, 2022.

[5] Ritesh Jha, Vandana Bhattacharjee, and Abhijit Mustafi. Increasing the prediction accuracy for thyroid disease: a step towards better health for society. *Wireless Personal Communications*, 122(2):1921–1938, 2022.

[6] Xiangdong Guo, Haifeng Zhao, and Zhenyu Tang. An improved deep learning approach for thyroid nodule diagnosis. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 296–299. IEEE, 2020.

[7] Dat Tien Nguyen, Jin Kyu Kang, Tuyen Danh Pham, Ganbayar Batchuluun, and Kang Ryoung Park. Ultrasound image-based diagnosis of malignant thyroid nodule using artificial intelligence. *Sensors*, 20(7):1822, 2020.

[8] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[9] Xueyan Mei, Xiaomeng Dong, Timothy Deyer, Jingyi Zeng, Theodore Trafalis, and Yan Fang. Thyroid nodule benignty prediction by deep feature extraction. In *2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 241–245. IEEE, 2017.

[10] Ye Zhu, Zhuang Fu, and Jian Fei. An image augmentation method using convolutional network for thyroid nodule classification by transfer learning. In *2017 3rd IEEE international conference on computer and communications (ICCC)*, pages 1819–1823. IEEE, 2017.