



Performance Evaluation Of Optimizers On Fine-Tuned VGG-19 Model For Thyroid Nodules Classification

Shiwangi kulhari^{1*}, Dr. Ritu vijay²

^{1*,2}*School of physical sciences Banasthali vidyapith Tonk, India*

**Corresponding Author: Shiwangi kulhari*

**School of physical sciences Banasthali vidyapith Tonk, India*

Abstract

The task of thyroid nodules classification involves the intricate analysis of patterns and features present in ultrasound images. Convolutional Neural Networks (CNNs) are being harnessed for thyroid malignancy detection due to their exceptional capability to process complex and high-dimensional medical imaging data effectively. This study introduces a meticulously fine-tuned VGG-19 CNN model, designed to cater specifically to the multi-classification of thyroid nodules within pre-processed ultrasound (US) images. Additionally, the model's efficacy is evaluated across a spectrum of optimization techniques. Experimental results underscore the model's effectiveness, showcasing accuracy rates of 0.6562, 0.8094, 0.8294, and 0.9201 when employing SGD, ADAGRAD, RMSprop, and ADAM optimizers, respectively, spanning 150 epochs. Importantly, the ADAM optimizer emerges as the key contributor to the optimal testing loss, signifying its crucial role in refining the model's performance.

CC License
CC-BY-NC-SA 4.0

Keywords— *thyroid nodules classification; fine-tuned; ultrasound; optimizers*

I. INTRODUCTION

Situated in the anterior neck, just beneath the Adam's apple, the thyroid gland is an essential endocrine organ. Its pivotal function involves synthesizing hormones through the utilization of iodine, thereby playing a vital role in regulating the body's hormonal balance [1][2]. Thyroid cancer frequently presents as an enlargement of the thyroid gland. Tumors that exhibit well-defined differentiation and encapsulation might be detectable through touch and could result in alterations in voice quality and challenges in breathing [3]. Thyroid cancer ranks as the most common type of endocrine malignancy, accounting for around 2.1% of all global cancer cases. The incidence of thyroid cancer has displayed a consistent increase in multiple countries over the past few decades [4]. Ultrasound (US) imaging is widely utilized as a diagnostic tool to detect and define the features of thyroid nodules. However, not only analyzing entire-slide images proves to be a challenging and labor-intensive task for thyroid specialists but also there is possibility of human error. [5][6]. Convolutional Neural Networks (CNNs) such as LeNet, AlexNet, NIN, ResNet, GoogLeNet, Xception and VGG [7] offer a solution to these challenges by applying deep techniques to ultrasound images. The learning performance of neural network model can be influenced on diverse widths and depths by utilizing different optimizers [8]. This study suggests employing a pre-trained VGG-19 model for fine-tuning and application of different optimizers such as SGD, ADAGRAD, RMSprop and ADAM to multi-classifying thyroid tumors due to VGG's

Available online at: <https://jazindia.com>

architectural simplicity and capacity to learn intricate image features. This paper illustrates the contribution of this task in this context-

- The pre-trained VGG19 convolutional neural network model is fine-tuned by adjusting filter counts and strategically introducing dropout layers with appropriate parameters after the convolutional layers.
- The fine-tuned VGG19 model is assessed by utilizing Optimizers such as SGD, Adagrad, RMSprop, and Adam over the course of 150 epochs.

The fine-tuned model demonstrated remarkable accuracy and exhibited noticeable variations in both accuracy and loss function values when different optimizers were employed.

II. LITERATURE SURVEY

Liang et al. have utilized a multi-organ Computer-Aided Diagnosis (CAD) system, employing convolutional neural networks (CNNs), to classify breast and thyroid nodules. They probed the effect of different preprocessing methods on diagnostic effectiveness [9]. Wang et al. performed a relative evaluation between radiomics and a deep learning-based approach, specifically the fine-tuned VGG-16, to classify thyroid nodules [10]. Xie et al. introduced an innovative structural design that integrates deep learning techniques with local binary pattern [11]. Liu et al. created a novel combined convolutional neural network (CNN) that incorporates information fusion techniques. [12]. Vadhiraaj et al. have performed a relative analysis between the support vector machine (SVM) and artificial neural network (ANN) classification algorithms [13]. Li et al. have built a holistic automated system for recognizing and classifying CT images of thyroid tumors [14].

Hang has introduced a method that combines conventional and deep features to form a unified feature domain, and compared ResNet18 (an 18-layer residual CNN) with Res-GAN. [15]. Peng et al. have created ThyNet model, which amalgamated ResNet, DenseNet, and ResNeXt architectures [16]. Qi et al. have created a comprehensive network model named Mask-RCNN18. This model incorporated the residual network (ResNet) and the feature pyramid network (FPN) for feature extraction, employed the region proposal network (RPN) for classification, and integrated bounding box (BB) regression for generating Regions of Interest (ROI) to detect the existence of significant extrathyroidal extension (ETE) in cases of thyroid cancer [17]. Liu et al. have created ThyNet-LNM, a dedicated deep-learning model for the evaluation of Lymph Node Metastasis (LNM) [18]. Ajilisa et al. combined inception modules with squeeze and excitation networks to improve the inception network's recognition accuracy. They also utilized multi-level transfer learning approach [19].

Yang et al. have assessed neural network variance and bias and validated findings with theoretical analysis of linear networks [20]. Keskar et al. have explored a hybrid approach that initiates training with an adaptive method and transitions to SGD when suitable. They additionally introduced SWATS, a straightforward strategy that shifts from Adam to SGD upon fulfilling a triggering condition [21]. Choi et al. have showcased how hyperparameter tuning protocol can influence the sensitivity of optimizer [22]. Nado et al. juxtaposed outcomes from LARS and LAMB against conventional optimization algorithms like ADAM and Nesterov momentum, focusing on large batch sizes [23].

III. MATERIALS AND METHODS

This part is organized into five primary segments: dataset, data preprocessing, classical VGG-19 Model, proposed fine-tuning strategy for the VGG-19 Model, and optimization techniques. The initial segment outlines the process of gathering and obtaining ultrasound images. The second part encompasses the procedures for preprocessing the images. In the third section, the classical VGG-19 model is expounded upon. The fourth segment elucidates the strategy employed to fine-tune the VGG-19 model. Lastly, the fifth section Offers a brief summary of the employed optimization techniques.

A. Dataset

In this study, the Digital Database of Thyroid Ultrasound Images (DDTI), an openly accessible dataset, is employed. This dataset comprises the assessment of 347 B-mode thyroid ultrasound images. These images were extracted from video sequences of thyroid ultrasounds captured using ultrasound devices, specifically TOSHIBA Nemio MX and TOSHIBA Nemio 30. A group of 299 patients with thyroid-related conditions had their images evaluated by two experienced radiologists. Each image is associated with a comprehensive

annotation and diagnostic description stored in an XML file. Radiologists employed the TIRADS (Thyroid Imaging Reporting and Data System) system for patient classification. TIRADS assigns points to different ultrasound features of a nodule, allocating additional points to more suspicious characteristics. The overall TI-RADS score is calculated by aggregating the feature points across all categories-TR1-normal thyroid; TR2- benign; TR3- absence of suspicious ultrasound features; TR4a- presence of a single suspicious ultrasound feature; TR4b- presence of 2 suspicious ultrasound features; TR4c- presence of 3 or 4 suspicious ultrasound features; TR5- presence of 5 suspicious ultrasound features. Descriptions corresponding to the images are also incorporated in the XML files. Within this dataset, the experts delineated the boundaries of the nodules and annotated distinctive features like veins, trachea, muscles, arteries, and calcifications. [24][25].

B. Data preprocessing

The images undergo cropping to specific dimensions, maintaining a square shape in the cropped region. Subsequently, these images are converted into a single channel to ensure data uniformity, reduce dimensionality, and facilitate grayscale-specific processing by transforming RGB images into grayscale format. A sequence of additional image processing operations, including thresholding, denoising, contour detection, and resizing, is applied to generate processed images suitable for subsequent analysis or use. The dataset is partitioned into three categories: a training set comprising 300 ultrasound images, a validation set consisting of 13 ultrasound images, and a test set containing 34 ultrasound images. This partitioning forms the basis for training, fine-tuning, and evaluating the proposed model. This data separation ensures the model's capacity to generalize effectively to data it hasn't encountered before.

C. Data preprocessing

The images undergo cropping to specific dimensions, maintaining a square shape in the cropped region. Subsequently, these images are converted into a single channel to ensure data uniformity, reduce dimensionality, and facilitate grayscale-specific processing by transforming RGB images into grayscale format. A sequence of additional image processing operations, including thresholding, denoising, contour detection, and resizing, is applied to generate processed images suitable for subsequent analysis or use. The dataset is partitioned into three categories: a training set comprising 300 ultrasound images, a validation set consisting of 13 ultrasound images, and a test set containing 34 ultrasound images. This partitioning forms the basis for training, fine-tuning, and evaluating the proposed model. This data separation ensures the model's capacity to generalize effectively to data it hasn't encountered before.

D. Classical VGG-19 Model

The VGG-19 architecture, initially introduced by Simonyan and Zisserman in 2014 [26], is a deep convolutional neural network encompassing a total of 19 layers. Its primary objective is to classify images into a thousand distinct object classes. This is accomplished through training VGG-19 with the ImageNet database, which comprises 1 million images spanning these thousand categories. This architecture has gained widespread recognition in image classification tasks due to its effective utilization of multiple 3×3 filters within every convolutional layer. Fig. 1 illustrates the VGG-19 architecture, featuring 16 convolutional layers for feature extraction, followed by three fully-connected layers dedicated to classification. The feature extraction layers are grouped into 5 blocks, each succeeded by a max-pooling layer. The model accepts 224×224 dimension image as its input, and the label of the object present in the image is revealed in the model's output [27].

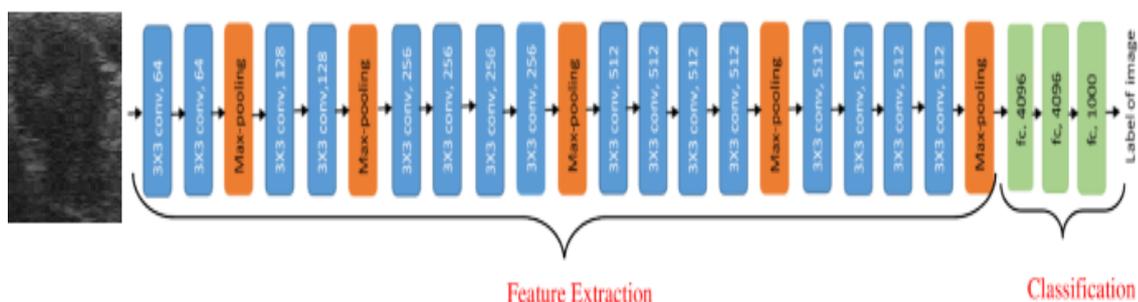


Fig.1. VGG-19 architecture [27]

E. Fine-tuning strategy for the VGG-19 Model

Fine-tuning of a CNN model entails modifying the parameters of a pre-trained neural network to align it with a particular task or dataset. This process enables the model to leverage the learned features from the initial dataset while accommodating the unique attributes of the new task. Furthermore, it accelerates training by utilizing pre-existing weights as a foundation. Utilizing a pre-trained model is especially advantageous when confronted with a limited thyroid malignancy dataset, as it brings along general knowledge from its original training.

The proposed fine-tuned model comprises 16 convolutional layers utilizing ReLU activation, along with 5 max pooling layers. Additionally, it integrates 2 fully-connected layers employing ReLU activation, culminating in an output fully-connected layer using softmax activation. 9 batch normalization layers are introduced to normalize activations within mini-batches. Moreover, 11 dropout layers are incorporated to strike a balance between accurately fitting the training data and effectively generalizing to unseen data.

In the process of fine-tuning, adjustments are made to the pre-trained VGG-19 model, including modifying the number of filters. Filters in early layers are designed to capture rudimentary features, whereas deeper layers capture more intricate features. This adjustment enables control over the trade-off between model complexity and computational efficiency. Convolutional layers employ the He Normal initialization method, which aids in achieving quicker and more stable convergence during training. This initialization method ensures initial weights are appropriately scaled to align with the characteristics of ReLU activations. 0.0001 is set as the learning rate. The categorical cross-entropy loss function is employed for the classification of thyroid nodules into multiple TIRADS categories, with the objective of minimizing loss and enhancing classification performance.

F. Optimization techniques

The model is compiled using four different optimizers one-by-one for comparison: SGD, Adagrad, RMSprop, and Adam.

- Stochastic gradient descent (SGD) [28] - It is a variant of the gradient descent optimization algorithm. Gradient descent is a technique used to reduce $J(\theta)$ an objective function, which depends on a model's parameters $\theta \in \mathbb{R}^d$. It achieves this by In the opposite direction of the gradient of the objective function denoted as $\nabla_{\theta}J(\theta)$, parameters are adjusted. η the learning rate, governs the scale of the steps taken to approach a (potentially local) minimum. For every training example $x^{(i)}$, a parameter update is executed, along with its corresponding label $y^{(i)}$ by Stochastic Gradient Descent as shown in equation 1:

$$\theta = \theta - \eta \cdot \nabla_{\theta}J(\theta; x^{(i)}; y^{(i)}) \quad (1)$$

- Adaptive Gradient Algorithm (ADAGrad) - Based on the historical gradient information, it adaptively scale the learning rate for each parameter. It achieves this by individually maintaining a summation of the gradients squared for every parameter. It addresses the challenge of selecting a suitable learning rate for different parameters in a model, especially when dealing with sparse data or varying gradients [29].
- Root Mean Square Propagation (RMSprop) [30] - It is an optimization algorithm devised to overcome certain limitations of conventional gradient descent optimization methods, like the issues of vanishing or exploding gradients. For each parameter the learning rate is dynamically adjusted during the training process. It is an enhancement of ADAGrad, specifically designed to enhance performance in situations involving non-convex optimization. It is grounded in the concept that normalizing the gradient vector by the root mean square value of every weight leads to improved learning outcomes.
- Adaptive Moment Estimation (ADAM) [31] - It is designed for the optimization of stochastic objective functions using first-order gradients. It achieves this by leveraging adaptive estimates of lower-order moments. Drawing inspiration from momentum, Adam utilizes the moving average of gradients and incorporates RMSprop's approach of squared gradient. This amalgamation renders Adam well-suited for addressing challenges presented by sparse and very noisy gradients.

IV. EXPERIMENTAL RESULTS

This part offers a thorough analysis of the conducted experiments and their resulting outcomes. It covers three essential aspects: the experimental setup, evaluation indexes, and the results and discussion

A. Experimental setup

A computer having Windows 10 operating system and equipped with a NVIDIA RTX A5000 graphics card, offering 24 gigabytes (GB) of GPU memory is used for the experiment. Python 3.10 and Keras 2.3.1, utilizing TensorFlow GPU 1.16 as the underlying framework are used for the deep learning environment. The model's training spanned 150 epochs, encompassing various stages of learning and refinement.

B. Evaluation Indexes

Accuracy is employed to assess the overall or average performance of a classification system. It refers to the closeness of measurement results to the true value. As presented in *equation (2)*, it measures how well the system correctly classifies instances in relation to the total count of instances.

$$Accuracy = \frac{(TNC+TPC)}{(TNC+TPC+FNC+FPC)} \quad (2)$$

Where, TNC=True negative cases tally; TPC= True positive cases tally; FNC=False negative cases tally; FPC=False positive cases tally

A graph depicting the training and testing loss is employed to serve as a tool for evaluating performance by assessing how well the model fits the data. It helps to determine whether the model is overfitting, underfitting, or achieving a balanced fit. This graph displays the change in loss values over the course of training and testing iterations, offering insights into how well the model is learning and generalizing.

C. Results and discussion

Table 1 presents a comparative analysis of the accuracy achieved by the fine-tuned VGG-19 model using four different optimization algorithms one-by-one - SGD, ADAgrad, RMSprop, and ADAM - across 150 epochs for the purpose of thyroid malignancy detection.

TABLE I.

S. No.	Optimizer	Accuracy
1.	SGD	0.6562
2.	ADAgrad	0.8094
3.	RMSprop	0.8294
4.	ADAM	0.9201

As shown in table 1, the observed accuracy trends among different optimization algorithms reveal insightful patterns. The initial utilization of the SGD optimizer resulted in a comparatively lower accuracy of 0.6562, indicating its limited efficacy in tailoring the model for the specified task. Subsequently, the ADAgrad optimizer exhibited enhancement, achieving an accuracy of 0.8094, underscoring the benefits of adaptive learning rates in outperforming the basic SGD approach. Further optimization through the RMSprop algorithm yielded an accuracy of 0.8294, emphasizing the advantageous adaptation of learning rates based on recent gradient history. Notably, the highest accuracy of 0.9201 was attained by employing the ADAM optimizer, showcasing the profound impact of its adaptive learning rates and incorporation of momentum. This achievement highlights ADAM's ability to navigate the optimization landscape more adeptly, resulting in markedly superior outcomes.

Figure 2 offers a comparative assessment of the training and testing loss curves resulting from the application of four distinct optimization algorithms - SGD, ADAgrad, RMSprop, and ADAM - in the fine-tuned VGG-19 model for the classification of thyroid nodules. Throughout the 150 epochs of training the fine-tuned VGG19 model with the SGD optimizer, a noticeable gap emerged between the training and testing losses, as depicted in Figure 2(a). The training loss, which reflects optimization progress, exhibited a consistent decline with each iteration. In contrast, the testing loss, which assesses the model's performance on unseen data, remained notably higher, suggesting the possibility of overfitting where the model performs well on training data but struggles with generalization.

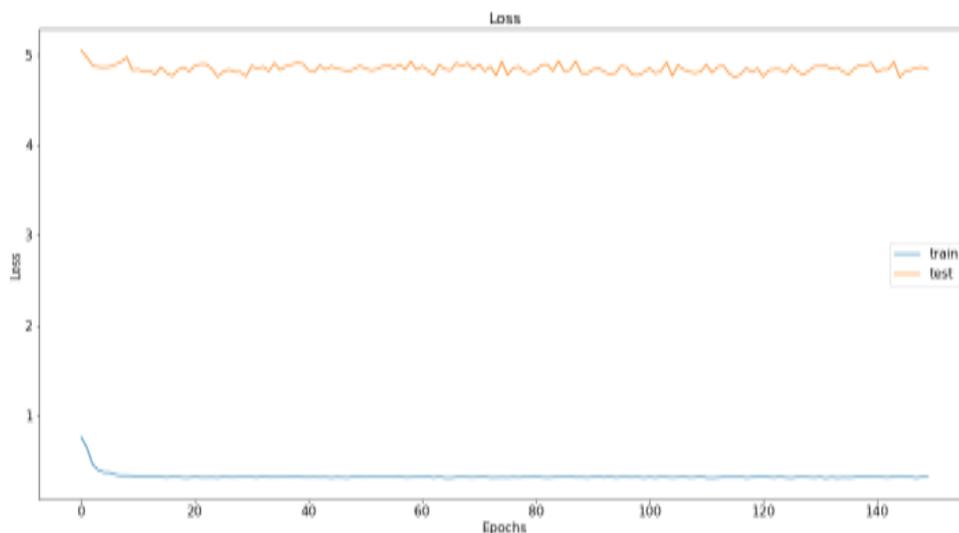
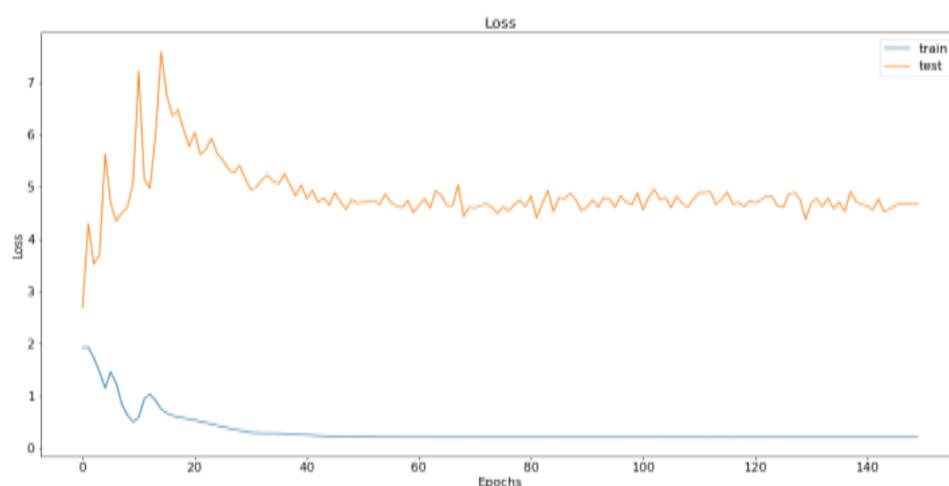
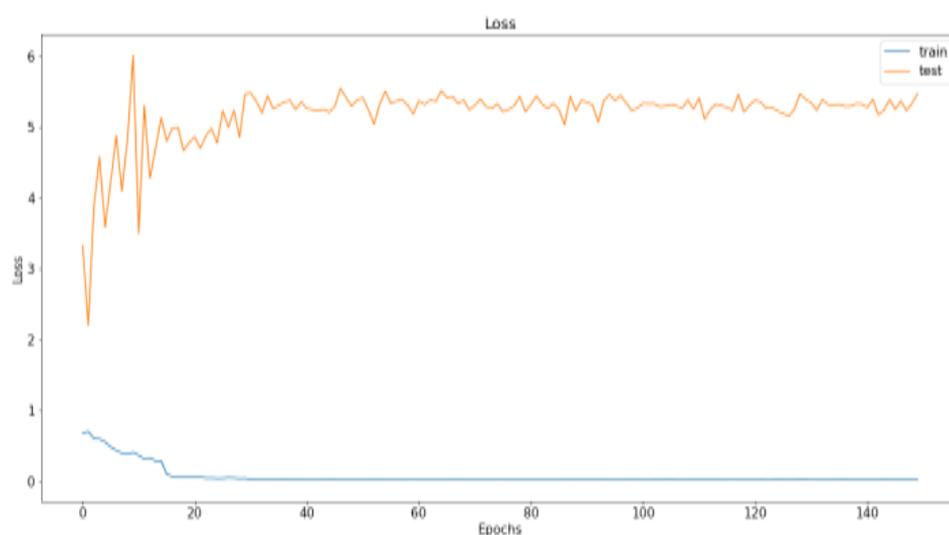


Fig. 2 (a) Training and Testing loss with SGD optimizer

For the models trained with the ADAGRAD optimizer (Figure 2(b)) and the RMSprop optimizer (Figure 2(c)), the training losses also decreased, albeit with initial fluctuations. However, the testing losses for these models stabilized at relatively high values over successive epochs, indicating challenges in achieving good generalization.

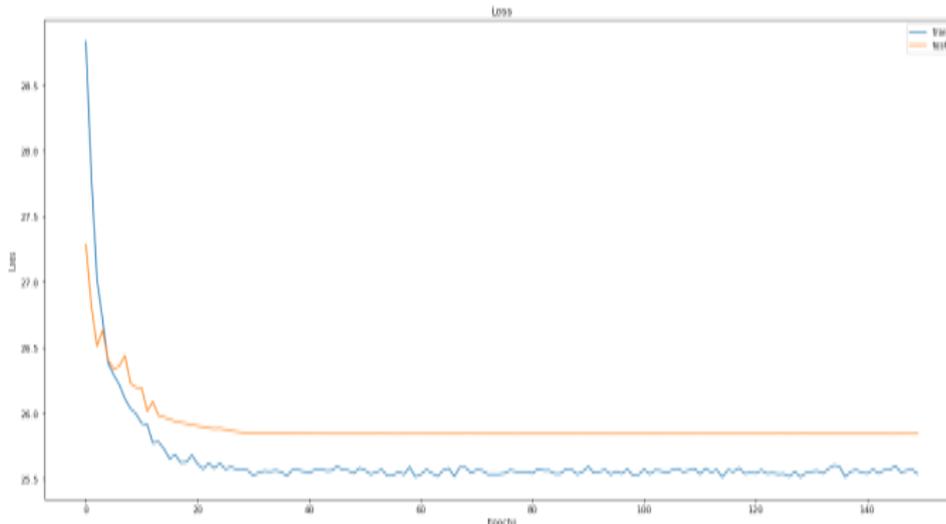


(b) Training and Testing loss with ADAGRAD optimizer



(c) Training and Testing loss with RMSprop optimizer

However, when the ADAM optimizer (Figure 2(d)) is utilized, both the training and testing loss exhibited significant reductions. Notably, there was a minimal disparity between the training and testing loss curves, suggesting a well-balanced fit. This subtle discrepancy signifies that the model achieved an optimal level of generalization, demonstrating the effectiveness of the ADAM optimizer in promoting both efficient learning and robust generalization



(d) Training and Testing loss with ADAM optimizer

V. CONCLUSION

In this paper, a fine-tuned VGG-19 model for the classification of thyroid nodules using ultrasound images is presented. The fine-tuning process not only shortened the training time of the model but also enhanced its capability to capture subtle characteristics specific to the medical domain. The model's effectiveness is scrutinized across various optimization techniques, and among them, ADAM optimizer stands out with the highest accuracy and the least testing loss, making it the most suitable optimizer for this particular task.

REFERENCES

1. Nguyen, Q. T., Lee, E. J., Huang, M. G., Park, Y. I., Khullar, A., & Plodkowski, R. A. (2015). Diagnosis and treatment of patients with thyroid cancer. *American health & drug benefits*, 8(1), 30–40.
2. Cancer of the Thyroid - Cancer Stat Facts. (n.d.). SEER. <https://seer.cancer.gov/statfacts/html/thyro.html>
3. Thyroid cancer - Symptoms and causes. (2023, April 26). Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/thyroid-cancer/symptoms-causes/syc-20354161>
4. Kitahara, C. M., Körmendiné Farkas, D., Jørgensen, J. O. L., Cronin-Fenton, D., & Sørensen, H. T. (2018, March 23). Benign Thyroid Diseases and Risk of Thyroid Cancer: A Nationwide Cohort Study. *The Journal of Clinical Endocrinology & Metabolism*, 103(6), 2216–2224. <https://doi.org/10.1210/jc.2017-02599>
5. Anari, S., Tataei Sarshar, N., Mahjouri, N., Dorosti, S., & Rezaie, A. (2022, August 25). Review of Deep Learning Approaches for Thyroid Cancer Diagnosis. *Mathematical Problems in Engineering*, 2022, 1–8. <https://doi.org/10.1155/2022/5052435>
6. Zhang, X., Lee, V. C. S., Rong, J., Liu, F., & Kong, H. (2022). Multi-channel convolutional neural network architectures for thyroid cancer detection. *PloS one*, 17(1), e0262128. <https://doi.org/10.1371/journal.pone.0262128>
7. Zhang, Q., Zhang, M., Chen, T., Sun, Z., Ma, Y., & Yu, B. (2019, January). Recent advances in convolutional neural network acceleration. *Neurocomputing*, 323, 37–51. <https://doi.org/10.1016/j.neucom.2018.09.038>
8. Dogo, E. M., Afolabi, O. J., & Twala, B. (2022, November 23). On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification. *Applied Sciences*, 12(23), 11976. <https://doi.org/10.3390/app122311976>

9. Liang, X., Yu, J., Liao, J., & Chen, Z. (2020, January 10). Convolutional Neural Network for Breast and Thyroid Nodules Diagnosis in Ultrasound Imaging. *BioMed Research International*, 2020, 1–9. <https://doi.org/10.1155/2020/1763803>
10. Wang, Y., Yue, W., Li, X., Liu, S., Guo, L., Xu, H., Zhang, H., & Yang, G. (2020). Comparison Study of Radiomics and Deep Learning-Based Methods for Thyroid Nodules Classification Using Ultrasound Images. *IEEE Access*, 8, 52010–52017. <https://doi.org/10.1109/access.2020.2980290>
11. Xie, J., Guo, L., Zhao, C., Li, X., Luo, Y., & Jianwei, L. (2020, December 1). A Hybrid Deep Learning and Handcrafted Features based Approach for Thyroid Nodule Classification in Ultrasound Images. *Journal of Physics: Conference Series*, 1693(1), 012160. <https://doi.org/10.1088/1742-6596/1693/1/012160>
12. Liu, Z., Zhong, S., Liu, Q., Xie, C., Dai, Y., Peng, C., Chen, X., & Zou, R. (2021, January 6). Thyroid nodule recognition using a joint convolutional neural network with information fusion of ultrasound images and radiofrequency data. *European Radiology*, 31(7), 5001–5011. <https://doi.org/10.1007/s00330-020-07585-z>
13. Vadhiraj, V. V., Simpkin, A., O’Connell, J., Singh Ospina, N., Maraka, S., & O’Keeffe, D. T. (2021, May 24). Ultrasound Image Classification of Thyroid Nodules Using Machine Learning Techniques. *Medicina*, 57(6), 527. <https://doi.org/10.3390/medicina57060527>
14. Li, W., Cheng, S., Qian, K., Yue, K., & Liu, H. (2021, May 27). Automatic Recognition and Classification System of Thyroid Nodules in CT Images Based on CNN . *Computational Intelligence and Neuroscience*, 2021, 1–11. <https://doi.org/10.1155/2021/5540186>
15. Hang, Y. (2021, July 22). Thyroid Nodule Classification in Ultrasound Images by Fusion of Conventional Features and Res-GAN Deep Features. *Journal of Healthcare Engineering*, 2021, 1–7. <https://doi.org/10.1155/2021/9917538>
16. Peng, S., Liu, Y., Lv, W., Liu, L., Zhou, Q., Yang, H., Ren, J., Liu, G., Wang, X., Zhang, X., Du, Q., Nie, F., Huang, G., Guo, Y., Li, J., Liang, J., Hu, H., Xiao, H., Liu, Z., . . . Xiao, H. (2021, April). Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *The Lancet Digital Health*, 3(4), e250–e259. [https://doi.org/10.1016/s2589-7500\(21\)00041-8](https://doi.org/10.1016/s2589-7500(21)00041-8)
17. Qi, Q., Huang, X., Zhang, Y., Cai, S., Liu, Z., Qiu, T., Cui, Z., Zhou, A., Yuan, X., Zhu, W., Min, X., Wu, Y., Wang, W., Zhang, C., & Xu, P. (2023, April). Ultrasound image-based deep learning to assist in diagnosing gross extrathyroidal extension thyroid cancer: a retrospective multicenter study. *EClinicalMedicine*, 60, 102007. <https://doi.org/10.1016/j.eclinm.2023.102007>
18. Liu, Y., Lai, F., Lin, B., Gu, Y., Chen, L., Chen, G., Xiao, H., Luo, S., Pang, Y., Xiong, D., Li, B., Peng, S., Lv, W., Alexander, E. K., & Xiao, H. (2023, June). Deep learning to predict cervical lymph node metastasis from intraoperative frozen section of tumour in papillary thyroid carcinoma: a multicentre diagnostic study. *E Clinical Medicine*, 60, 102007. <https://doi.org/10.1016/j.eclinm.2023.102007>
19. Ajilisa, O. A., Jagathy Raj, V. P., & Sabu, M. K. (2023, July 24). A Deep Learning Framework for the Characterization of Thyroid Nodules from Ultrasound Images Using Improved Inception Network and Multi-Level Transfer Learning. *Diagnostics*, 13(14), 2463. <https://doi.org/10.3390/diagnostics13142463>
20. Yang, Z., Yu, Y., You, C., Steinhardt, J., & Ma, Y. (2020, February 26). Rethinking Bias-Variance Trade-off for Generalization of Neural Networks. *arXiv.org*. <https://arxiv.org/abs/2002.11328v3>
21. Keskar, N. S., & Socher, R. (2017, December 20). Improving Generalization Performance by Switching from Adam to SGD. *arXiv.org*. <https://arxiv.org/abs/1712.07628v1>
22. Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C. J., & Dahl, G. E. (2019, October 11). On Empirical Comparisons of Optimizers for Deep Learning. *arXiv.org*. <https://arxiv.org/abs/1910.05446v3>
23. Nado, Z., Gilmer, J. M., Shallue, C. J., Anil, R., & Dahl, G. E. (2021, February 12). A Large Batch Optimizer Reality Check: Traditional, Generic Optimizers Suffice Across Batch Sizes. *arXiv.org*. <https://arxiv.org/abs/2102.06356v3>
24. Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., & Romero, E. (2015, January 28). An open access thyroid ultrasound image database. *SPIE Proceedings*. <https://doi.org/10.1117/12.2073532>
25. Song, R., Zhang, L., Zhu, C., Liu, J., Yang, J., & Zhang, T. (2020). Thyroid Nodule Ultrasound Image Classification Through Hybrid Feature Cropping Network. *IEEE Access*, 8, 64064–64074. <https://doi.org/10.1109/access.2020.2982767>
26. Simonyan, K., & Zisserman, A. (2014, September 4). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv.org*. <https://arxiv.org/abs/1409.1556v6>

27. Bansal, M., Kumar, M., Sachdeva, M., & Mittal, A. (2021, September 17). Transfer learning for image classification using VGG19: Caltech-101 image data set. *Journal of Ambient Intelligence and Humanized Computing*, 14(4), 3609–3620. <https://doi.org/10.1007/s12652-021-03488-z>
28. Ruder, S. (2016, September 15). An overview of gradient descent optimization algorithms. arXiv.org. <https://arxiv.org/abs/1609.04747v2>
29. Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7).
30. Dogo, E. M., Afolabi, O. J., & Twala, B. (2022, November 23). On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification. *Applied Sciences*, 12(23), 11976. <https://doi.org/10.3390/app122311976>
31. Kingma, D. P., & Ba, J. (2014, December 22). Adam: A Method for Stochastic Optimization. arXiv.org. <https://arxiv.org/abs/1412.6980v9>