

OPTIMIZING ULTRASOUND IMAGE CLASSIFICATION THROUGH TRANSFER LEARNING: FINE-TUNING STRATEGIES AND CLASSIFIER IMPACT ON PRE-TRAINED INNER-LAYERS

Mohamed Bal-Ghaoui, My Hachem El Yousfi Alaoui, Abdelilah Jilbab, Abdennaser Bourouhou

Mohammed V University in Rabat, National High School of Arts and Crafts, Electrical Engineering Department, E2SN Research Laboratory, Rabat, Morocco

Abstract. Transfer Learning (TL) is a popular deep learning technique used in medical image analysis, especially when data is limited. It leverages pre-trained knowledge from State-Of-The-Art (SOTA) models and applies it to specific applications through Fine-Tuning (FT). However, fine-tuning large models can be time-consuming, and determining which layers to use can be challenging. This study explores different fine-tuning strategies for five SOTA models (VGG16, VGG19, ResNet50, ResNet101, and InceptionV3) pre-trained on ImageNet. It also investigates the impact of the classifier by using a linear SVM for classification. The experiments are performed on four open-access ultrasound datasets related to breast cancer, thyroid nodules cancer, and salivary glands cancer. Results are evaluated using a five-fold stratified cross-validation technique, and metrics like accuracy, precision, and recall are computed. The findings show that fine-tuning 15% of the last layers in ResNet50 and InceptionV3 achieves good results. Using SVM for classification further improves overall performance by 6% for the two best-performing models. This research provides insights into fine-tuning strategies and the importance of the classifier in transfer learning for ultrasound image classification.

Keywords: CNN, transfer learning, fine-tuning, SVM, ultrasound images, cancer classification

OPTIMALIZACJA KLASYFIKACJI OBRAZÓW ULTRASONOGRAFICZNYCH TECHNIKĄ TRANSFER LEARNING: STRATEGIE DOSTRAJANIA I WPLYW KLASYFIKATORA NA WSTĘPNIE WYTRENOWANE WARSTWY WEWNĘTRZNE

Streszczenie. Transfer Learning (TL) to popularna technika głębokiego uczenia stosowana w analizie obrazów medycznych, zwłaszcza gdy ilość danych jest ograniczona. Wykorzystuje ona wstępnie wyszkoloną wiedzę z modeli State-Of-The-Art (SOTA) i zastosowanie ich do konkretnych aplikacji poprzez dostrajanie (Fine-Tuning – FT). Jednak dostrajanie dużych modeli może być czasochłonne, a określenie, których warstw użyć, może stanowić wyzwanie. W niniejszym badaniu przeanalizowano różne strategie dostrajania dla pięciu modeli SOTA (VGG16, VGG19, ResNet50, ResNet101 i InceptionV3) wstępnie wytrenowanych na ImageNet. Zbadano również wpływ klasyfikatora przy użyciu liniowej SVM do klasyfikacji. Eksperymenty przeprowadzono na czterech ogólnodostępnych zbiorach danych ultrasonograficznych związanych z rakiem piersi, rakiem guzków tarczycy i rakiem gruczołów ślinowych. Wyniki są oceniane przy użyciu techniki pięciowarstwowej walidacji krzyżowej, a wskaźniki takie jak dokładność, precyzja i odzyskiwanie są obliczane. Wyniki pokazują, że dostrajenie 15% ostatnich warstw w ResNet50 i InceptionV3 osiąga dobre wyniki. Użycie SVM do klasyfikacji dodatkowo poprawia ogólną wydajność o 6% dla dwóch najlepszych modeli. Badania te zapewniają informacje na temat strategii dostrajania i znaczenia klasyfikatora w uczeniu transferowym dla klasyfikacji obrazów ultrasonograficznych.

Słowa kluczowe: CNN, transfer learning, dostrajanie, SVM, obrazy ultrasonograficzne, klasyfikacja nowotworów

Introduction

Medical ultrasound imaging is a widely used modality for diagnosing various conditions, such as tumors, cysts, and abnormalities in organs and tissues. It is a non-invasive technique, less expensive, and can provide real-time images for diagnosis purposes [15].

Accurate classification of ultrasound medical images plays a crucial role in the clinical decision-making process. However, it can be challenging due to the complexity of these images, as well as the limited availability of annotated data for training Deep Learning (DL) models based on Convolutional Neural Networks (CNNs) [7, 21]. These kinds of models have shown remarkable success in image classification tasks, but they may face limitations in medical ultrasound images due to the previous-mentioned obstacles.

In recent years, Transfer Learning (TL) has shown promising results in various computer vision tasks, and become a prominent DL technique allowing models trained on large datasets such as ImageNet, to be fine-tuned on smaller target datasets. TL has the potential to address the data deficiency limitations found in medical images, making it a valuable tool for improving the accuracy and efficiency of Computer-Aided-Diagnosis (CAD) systems.

TL pre-trained models have already learned low-level and generic features common to many images such as edges, contours, shapes, and so on. While high-level features specific to the classification task are learned through the classifier by means of Fine-Tuning (FT). This technique can help reduce the computational cost of DL models, saving the effort of building layers from scratch. This can be especially beneficial in the medical field where annotated data are scarce and costly to obtain.

Despite its advantages, TL requires essentially a classifier built on-top for FT purposes. The classifier can be as simple as a Global Max-Pooling (GMP) or Flatten layer with Fully Connected (FC) layers that match the number of target classes. Utilizing such a shallow Multilayer Perceptron (MLP) classifier may not guarantee optimal results and can, in certain cases, result in overfitting the training set due to the depth of the pre-trained models. However, choosing an adequate FT strategy and a questionable classifier can help overcome these limitations. In such cases, hybrid approaches combining CNNs extracted features along with Machine Learning (ML) estimators can offer improved performance and robustness.

In recent years, there has been growing interest in combining CNNs with other classifiers to improve accuracy, robustness, and interpretability. One popular hybrid approach is the combination of CNN and Support Vector Machines (SVM) [4].

CNNs are known for their ability to learn hierarchical features from images, automatically capturing relevant patterns and structures. SVM, on the other hand, is a well-known ML estimator that was initially designed for binary classification tasks providing a clear decision boundary while handling small datasets effectively. Combining the feature extraction capabilities of CNNs with the discriminative power of SVM, this approach can lead to a better-generalized classification outcome.

In this paper, we emphasize the significance of FT techniques in TL for the classification of ultrasound images. We evaluate five of the classically recognized ImageNet pre-trained models, namely VGG16, VGG19, ResNet50, ResNet101, and InceptionV3, using various FT strategies. Additionally, we investigate the effectiveness of the hybrid CNN-SVM approach by incorporating an SVM classifier on top of the fine-tuned pre-trained models. The findings of this study provide valuable insights into the advantages of fine-tuning specific layers and the hybrid CNN-SVM approach for ultrasound image classification.

1. Related works

In recent research, TL has been widely used to address the demand for large labeled data required to train DL models. Training these models from scratch for medical images can be challenging due to several reasons. Notably, dataset availability is often expensive in collection and storage, particularly when involving professional radiologists in the annotation process which is time-consuming and error-prone posing challenges in training DL models. As a result, researchers have introduced TL as an efficient and low-cost technique to remediate the lack of data. Many researchers have used ImageNet pre-trained models, such as VGG16, ResNet, and InceptionV3 in various medical imaging applications, such as skin cancer, breast cancer, and so on.

TL commonly employs two strategies: fine-tuning, utilized when the target dataset is enough for the training, and feature extraction, an alternative method for leveraging low-level features from pre-trained models. Besides, data augmentation techniques such as rotation, cropping, noise adding, and color manipulation

have also been commonly used to expand the dataset and prevent overfitting [8, 21].

TL has made major contributions to medical image analysis by overcoming the problem of data scarcity and saving time and hardware resources. In the review paper [5], authors investigated 121 studies around selecting the best backbone models and TL approaches for medical images. They have therefore divided TL strategies into four categories including feature extractor, feature extractor hybrid, fine-tuning, and fine-tuning from scratch. Authors have also declared Inception, ResNet, VGG, AlexNet, and LeNet as the most common TL models used in literature. The same was also confirmed in the review paper [6]. Additionally, authors of [5] have recommended preferring ResNet and Inception models as feature extractors due to their performance and computational efficiency.

In the same context, authors in [14] suggest a novel deep learning cascaded feature framework to address the issue of the high dimensionality of features extracted from deep layers of pre-trained CNNs models. Their framework utilized pre-trained models such as AlexNet, VGG, and GoogleNet to extract shallow and deep features, and have employed their univariate strategy to overcome the dimensionality and multicollinearity issues in the extracted features. The evaluation of their proposed framework yielded an accuracy of 98.50%, sensitivity of 98.06%, specificity of 98.99%, and precision of 98.98%.

The authors of the paper [3], investigated ten TL models as backbones for the U-Net [12] model for segmenting breast ultrasound images. The obtained results demonstrated the efficiency of pre-trained models in extracting relevant features for breast lesions segmentation.

In [19], authors used deep learning pre-trained models such as ResNet50, DenseNet121, and EfficientNetB3, besides transformer-based methods such as ViT-B/16 to classify Salivary gland tumors. These tumors are commonly inferred from the parotid glands where only 20% of the tumors are malignant. Authors have performed a binary classification on a dataset of 251 patients, with about 29.5% of the malignant cases. They used data augmentation techniques during the training such as random flipping, rotating, blurring, and lighting adjustments. Their results outperformed those of inexperienced radiologists. Notably, EfficientNetB3 and DenseNet121 models achieved accuracy and Area Under the Curve (AUC) of 80%, 0.82, and 77%, 0.81 respectively. In the same context, [20] have trained a modified ResNet18 network over 1200 epochs with a learning rate of 1e-6 using Adam optimizer on a dataset of parotid lesions from 232 patients and a total of 3791 cropped parotid gland region images. The dataset was partitioned into 90% of training and 10% of validation sets. Additionally, data augmentation techniques such as image flipping and contrast adjustment were used for data enhancement. Authors have reported an accuracy of 82.18% with a micro-AUC of 0.93.

In the context of hybrid models, authors in [1] and [17] demonstrated the advantages of using linear L2-SVM as a top layer instead of softmax in DL architectures, highlighting benefits such as differentiability and stronger error penalization. They showed that L2-SVM is slightly better than L1-SVM, and they used linear SVMs in their experiments. They tested their approach on well-known datasets and achieved competitive results in a facial expression recognition competition. They highlighted the effectiveness of the last layer SVM in comparison to softmax, and they attributed the performance gain to the superior regularization effects of the SVM loss function rather than better parameter optimization.

The work in [9] develops a generic CAD system based on features extracted from pre-trained CNNs tested on 12 open-access image datasets. The authors aimed to explore the power of intermediate and last layers of ImageNet pre-trained models such as GoogleNet (Inception), ResNet, and DenseNet201 feature vectors for training an ensemble of SVMs. The extracted features are fed to SVMs, then combined for the final results.

Regarding thyroid nodule classification, [16] suggested a hybrid model based on CNN and SVM, compiled with a hinge loss function. They evaluated the results on two public datasets containing 1180 and 2616 thyroid ultrasound images after applying data augmentation. They reported an accuracy of 94.57%, 96% and specificity of 91.89%, 93.93%, and a sensitivity of 96.70%, 97.80% for both datasets, dataset-1 and dataset-2 respectively.

To guarantee optimal performances, SVM hyperparameters such as the C penalty parameter, and the kernel were investigated. [18] used Quantum-Behaved Particle Swarm Optimization (QPSO) algorithm to optimize SVM parameters due to its global search ability and fewer control parameters. A hybrid model consisting of a LeNet-5 network and SVM was utilized and validated on breast cancer cell images. The model achieved a test accuracy of 93.15%, outperforming the model without SVM, by 1.9%.

2. Proposed method

In this section, we introduce and discuss the different phases involved in this study, with a particular emphasis on FT strategies. Then, we delve into the details of the FT hybrid CNN-SVM approach and explore its potential in enhancing the overall classification performances. The study workflow is described in Fig. 1.

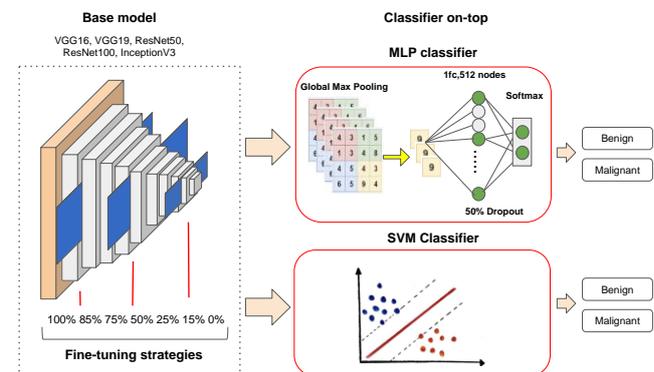


Fig. 1. Study workflow

2.1. Study workflow

To evaluate the experiment, four publicly available ultrasound datasets were used representing diverse anatomical regions and clinical cancer scenarios. Breast, thyroid nodules, and salivary glands are the organs investigated with their corresponding datasets 1) Breast Mendeley [11], 2) Breast Ultrasound Images (BUSI) [2], 3) Digital Database of Thyroid Images (DDTI) [10], and 4) Salivary Glands from Ultrasound cases website [13].

TL FT strategies were separately applied to each dataset using four state-of-the-art pre-trained models: VGG16, VGG19, ResNet50, ResNet101, and InceptionV3.

A model selection mechanism was implemented using a 5-fold Stratified Cross-Validation (SCV) technique, where the best model was saved during each fold for subsequent evaluation. This technique is widely recognized and effective in handling imbalanced datasets, which is often the case for medical images. Additionally, classification evaluation metrics including accuracy, precision, recall, and AUC values were computed. Furthermore, the Receiver Operator Characteristic (RoC) curve and Confusion Matrix (CM) were provided for a comprehensive description of the classification outcomes for the two best-performing models.

FT strategies and the effectiveness of the hybrid CNN-SVM approach were separately evaluated for each dataset. This enabled a thorough evaluation of the proposed approaches within the specific medical ultrasound domain, offering valuable insights into their performance and applicability. Figure 2 displays four samples from each dataset, while table 1 provides an overview of their class distribution.

2.2. Models configurations

A. Phase 1: Transfer learning fine-tuning approach (FT)

This approach aims at extracting optimal features from ultrasound images. For this task, pre-trained models with ImageNet weights were used as backbones, followed by a shallow MLP classifier with a GlobalMaxPooling layer, a fully connected

layer (512 nodes) with 50% dropout, and a final fully connected layer (2 nodes) with softmax activation function.

Table 2 describes the seven fine-tuning strategies that were implemented, irrespective of the models' depth, and highlights the number of total and frozen layers for the investigated pre-trained models.

Models in this phase were compiled using binary cross-entropy and Adam optimizer, then trained for 30 epochs with a learning rate of 0.0001 and a batch size of 4. Models evaluation was conducted using a Stratified Cross-Validation technique with metrics computed and averaged across the 5-folds representing the overall model performance.

B. Phase 2: Transfer learning fine-tuning hybrid approach (FT-SVM)

This approach consisted of employing SVM in replacement of the basic classifier made earlier. SVM inputs were obtained from the best features extracted from the FT models that were saved during the earlier phase.

Based on the literature review, the linear kernel-based SVM classifier has proved its efficiency in delivering good results in image classification tasks. Therefore, only SVM's C-hyperparameter was tuned using a GridSearch mechanism with a 5-fold SCV technique for values ranging from 0.1, 1, 10, and 100. SVM's C parameter is known to penalize each misclassified point and controls the SVM decision boundaries. Thus, choosing the right C value may require more thorough tests, hence the used GridSearch mechanism.

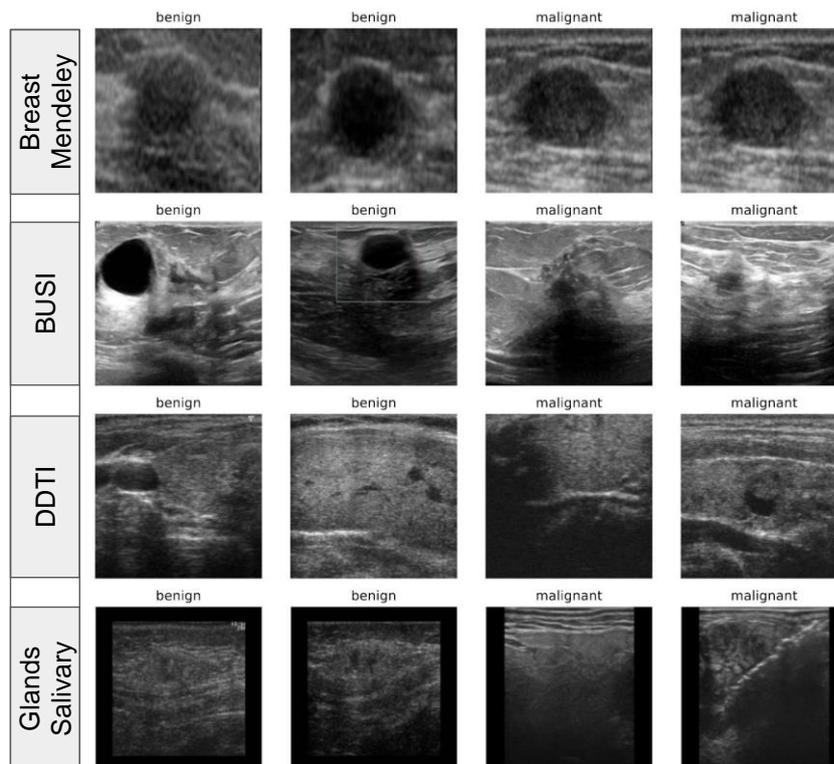


Fig. 2. Ultrasound datasets overview

Table 1. Datasets class distribution

Dataset	Benign	Malignant	Total
Breast Mendeley	100	150	250
Breast BUSI	437	210	647
Thyroid DDTI	49	399	448
Salivary Glands	781	109	890

Table 2. Fine-tuning strategies and pre-trained models: layers description

	VGG16	VGG19	ResNet50	ResNet101	InceptionV3
Total layers	19	22	175	345	311
FT strategies (%)	Number of frozen layers				
0% (purely TL)	19	22	175	345	311
15%	16	19	149	293	264
25%	14	16	131	259	233
50%	10	11	88	173	156
75%	5	6	44	86	78
85%	3	4	26	52	47
100% (scratch fine-tuning)	0	0	0	0	0

3. Experimental results and discussion

The investigated fine-tuning strategies yielded the results described in table 3. Since there are many metrics involved, we have structured the results in this table by the AUC value which reflects the true positive and the false negative rates of the classification. The AUC values were computed for each strategy, both for the fine-tuned models using a basic MLP classifier and the fine-tuned models using SVM. The best-performing models were chosen based on their AUC values performances across all models. Fine-tuning 15% of layers was found the best strategy and resulted in a good classification, specifically when performed by SVM. Table 4 shows the computed metrics for the five models along with the optimal SVM-C parameter value. Results in this table are structured as follows: each metric consists of a pair of values, representing the performance of the fine-tuned layers with the basic classifier and the fine-tuned layers with SVM.

For each dataset, excluding Breast Mendeley due to its potentially optimistic results that could impact the overall accuracy of the study, the confusion matrix and RoC curves were generated separately for the winning strategy in both ResNet50 and InceptionV3. Additionally, in the breast BUSI dataset, ResNet50 showed an increase of 18% in sensitivity, while InceptionV3 exhibited an increase of 12%. For the thyroid DDTI dataset, ResNet50 achieved a sensitivity increase of 3%, whereas InceptionV3 showed a 2% increase. In the case of the salivary glands dataset, ResNet50 demonstrated an increase of 0.52% in sensitivity compared to InceptionV3 with a 0.42%. Besides, InceptionV3 was found the most consistent model across all strategies and different ultrasound datasets being used. Figures 3, 5, and 7 display the confusion matrix of ResNet50 and InceptionV3 for the breast BUSI, thyroid DDTI, and Salivary Glands datasets, respectively, using the 15% FT strategy. Additionally, Figures 4, 6 and 8 illustrate the differences in RoC curves and emphasize the superior performance of SVM in the overall classification.

Table 3. Fine-tuning strategies by AUC values: FT and FT-SVM

AUC (%)	VGG16	VGG19	ResNet50	ResNet101	InceptionV3
Breast Mendeley	0% :(0.97, 0.95)	0% :(0.94, 0.97)	0% :(0.53, 0.70)	0% :(0.74, 0.97)	0% :(0.99, 1.00)
	15% :(0.94, 1.00)	15% :(0.76, 0.90)	15% :(1.00, 1.00)	15% :(1.00, 1.00)	15% :(1.00, 1.00)
	25% :(0.86, 1.00)	25% :(0.90, 1.00)	25% :(1.00, 1.00)	25% :(1.00, 1.00)	25% :(1.00, 1.00)
	50% :(1.00, 1.00)	50% :(0.94, 1.00)	50% :(0.99, 1.00)	50% :(0.95, 0.95)	50% :(1.00, 1.00)
	75% :(1.00, 1.00)	75% :(1.00, 0.98)	75% :(0.66, 0.97)	75% :(0.95, 1.00)	75% :(1.00, 1.00)
	85% :(1.00, 0.98)	85% :(1.00, 0.99)	85% :(0.66, 0.99)	85% :(0.97, 1.00)	85% :(1.00, 0.99)
Breast BUSI	100%:(1.00, 1.00)	100%:(0.97, 1.00)	100%:(1.00, 1.00)	100%:(1.00, 1.00)	100%:(1.00, 1.00)
	0% :(0.80, 0.81)	0% :(0.78, 0.84)	0% :(0.51, 0.78)	0% :(0.56, 0.76)	0% :(0.81, 0.84)
	15% :(0.76, 0.93)	15% :(0.67, 0.85)	15% :(0.87, 0.98)	15% :(0.81, 0.95)	15% :(0.90, 0.99)
	25% :(0.80, 0.90)	25% :(0.79, 0.88)	25% :(0.85, 0.93)	25% :(0.82, 0.95)	25% :(0.90, 0.98)
	50% :(0.84, 0.94)	50% :(0.78, 0.96)	50% :(0.81, 0.95)	50% :(0.75, 0.81)	50% :(0.88, 0.98)
	75% :(0.86, 0.96)	75% :(0.84, 0.95)	75% :(0.76, 0.89)	75% :(0.76, 0.83)	75% :(0.83, 0.97)
Thyroid DDTI	85% :(0.87, 0.95)	85% :(0.86, 0.94)	85% :(0.80, 0.90)	85% :(0.78, 0.85)	85% :(0.85, 0.96)
	100%:(0.64, 0.88)	100%:(0.60, 0.87)	100%:(0.90, 0.99)	100%:(0.84, 0.97)	100%:(0.88, 0.98)
	0% :(0.51, 0.50)	0% :(0.50, 0.50)	0% :(0.50, 0.50)	0% :(0.50, 0.50)	0% :(0.52, 0.50)
	15% :(0.50, 0.50)	15% :(0.50, 0.50)	15% :(0.59, 0.94)	15% :(0.49, 0.50)	15% :(0.54, 0.91)
	25% :(0.50, 0.50)	25% :(0.50, 0.50)	25% :(0.52, 0.83)	25% :(0.50, 0.50)	25% :(0.55, 0.92)
	50% :(0.52, 0.50)	50% :(0.50, 0.50)	50% :(0.50, 0.50)	50% :(0.50, 0.50)	50% :(0.55, 0.50)
Salivary Glands	75% :(0.50, 0.50)	75% :(0.50, 0.50)	75% :(0.50, 0.50)	75% :(0.50, 0.50)	75% :(0.51, 0.66)
	85% :(0.50, 0.55)	85% :(0.51, 0.59)	85% :(0.50, 0.51)	85% :(0.50, 0.50)	85% :(0.53, 0.62)
	100%:(0.50, 0.50)	100%:(0.50, 0.50)	100%:(0.55, 0.50)	100%:(0.54, 0.57)	100%:(0.51, 0.93)
	0% :(0.61, 0.50)	0% :(0.54, 0.50)	0% :(0.50, 0.50)	0% :(0.66, 0.50)	0% :(0.58, 0.50)
	15% :(0.50, 0.50)	15% :(0.50, 0.50)	15% :(0.67, 0.94)	15% :(0.61, 0.93)	15% :(0.71, 0.95)
	25% :(0.56, 0.75)	25% :(0.50, 0.50)	25% :(0.60, 0.61)	25% :(0.52, 0.50)	25% :(0.74, 0.90)

Table 4. Fine-tuning results for the 15% layers

Dataset	Metrics(%)	VGG16	VGG19	ResNet50	ResNet101	InceptionV3
Breast Mendeley	SVM-C parameter	1	0.1	0.1	0.1	1
	Accuracy	(0.9480, 1.0000)	0.7960, 0.9160)	(1.0000, 1.0000)	(1.0000, 1.0000)	(1.0000, 1.0000)
	Precision	(0.9419, 1.0000)	(0.7619, 0.8778)	(1.0000, 1.0000)	(1.0000, 1.0000)	(1.0000, 1.0000)
	Sensitivity	(0.9733, 1.0000)	(0.9600, 1.0000)	(1.0000, 1.0000)	(1.0000, 1.0000)	(1.0000, 1.0000)
Breast BUSI	SVM-C parameter	100	100	1	0.1	10
	Accuracy	(0.8176, 0.9412)	(0.7650, 0.8654)	(0.8887, 0.9768)	(0.8469, 0.9644)	(0.9072, 0.9876)
	Precision	(0.7875, 0.9234)	(0.7735, 0.7996)	(0.8415, 0.9494)	(0.8100, 0.9673)	(0.8504, 0.9764)
	Sensitivity	(0.6000, 0.8952)	(0.3904, 0.7904)	(0.8095, 0.9809)	(0.6904, 0.9238)	(0.8666, 0.9857)
Thyroid DDTI	SVM-C parameter	0.1	0.1	10	0.1	100
	Accuracy	(0.8893, 0.8906)	(0.8893, 0.8906)	(0.8849, 0.9821)	(0.8783, 0.8906)	(0.8783, 0.9753)
	Precision	(0.8893, 0.8906)	(0.8893, 0.8906)	(0.9088, 0.9852)	(0.8881, 0.8906)	(0.8970, 0.9778)
	Sensitivity	1.0000, 1.0000)	(1.0000, 1.0000)	(0.9676, 0.9949)	(0.9875, 1.0000)	(0.9751, 0.9949)
Salivary Glands	SVM-C parameter	0.1	0.1	10	1	10
	Accuracy	(0.8786, 0.8775)	(0.8775, 0.8775)	(0.8932, 0.9775)	(0.8853, 0.9662)	(0.8876, 0.9752)
	Precision	(1.0000, 0.0000)	(0.0000, 0.0000)	(0.6060, 0.9345)	(0.5744, 0.8545)	(0.5463, 0.8904)
	Sensitivity	(0.0091, 0.0000)	(0.0000, 0.0000)	(0.3669, 0.8809)	(0.2477, 0.8722)	(0.4862, 0.9090)

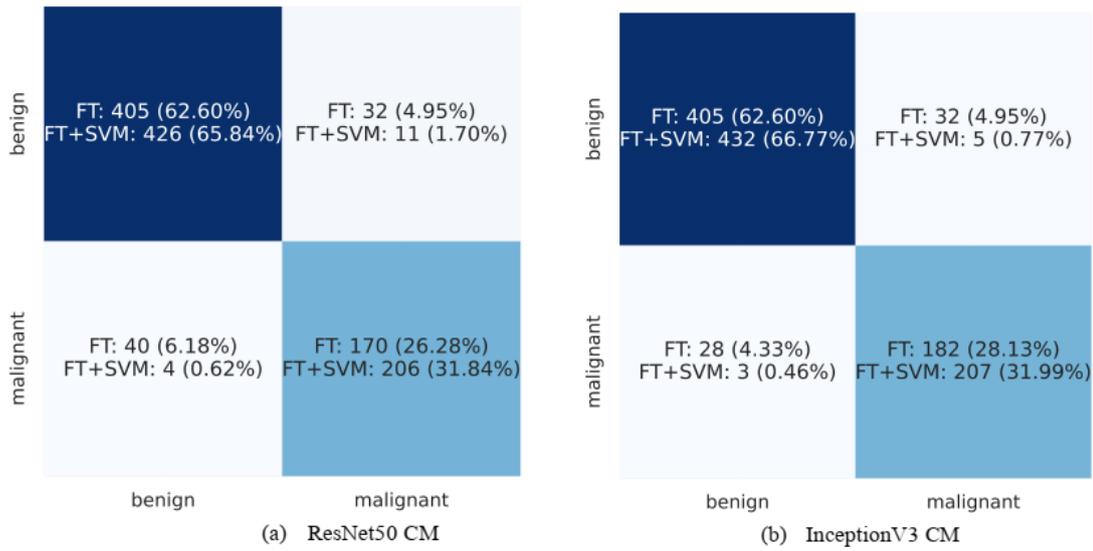


Fig. 3. BUSI confusion matrix curves: 15% of layers FT and FT+SVM

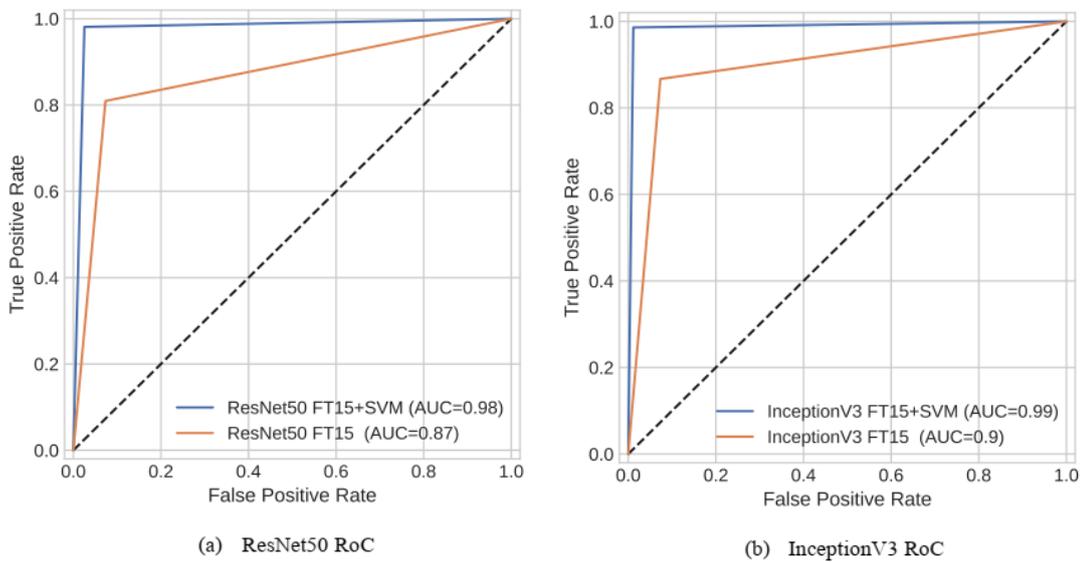


Fig. 4. BUSI RoC curve: 15% of layers FT and FT+SVM

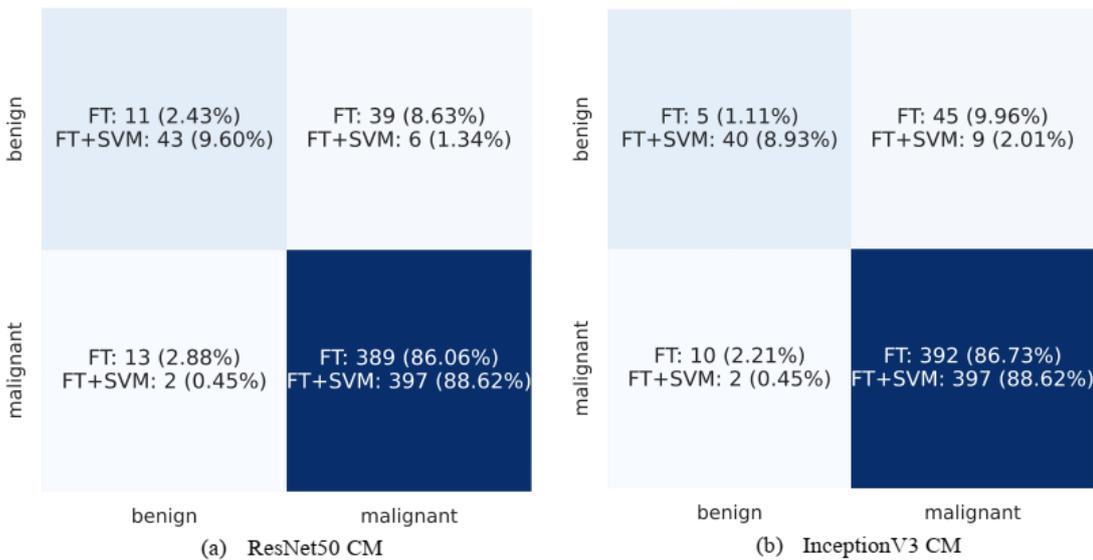
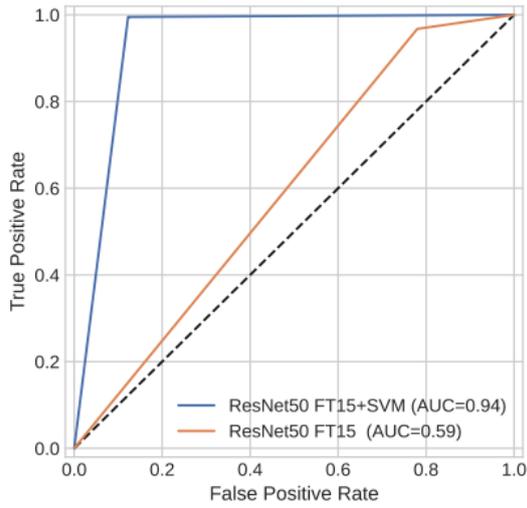
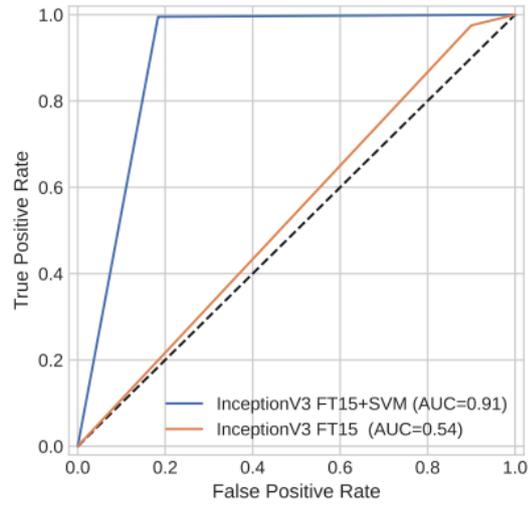


Fig. 5. DDTI confusion matrix: 15% of layers FT and FT+SVM

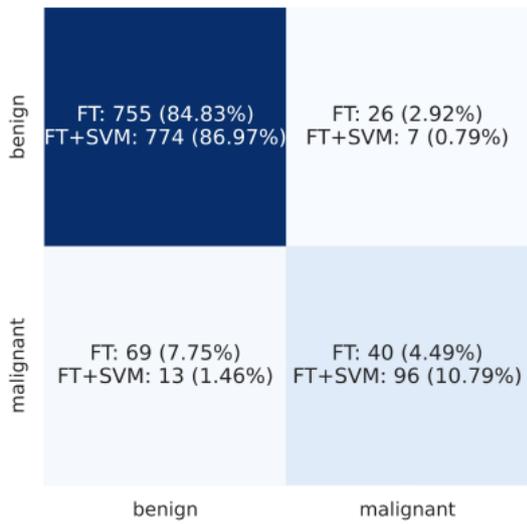


(a) ResNet50 RoC

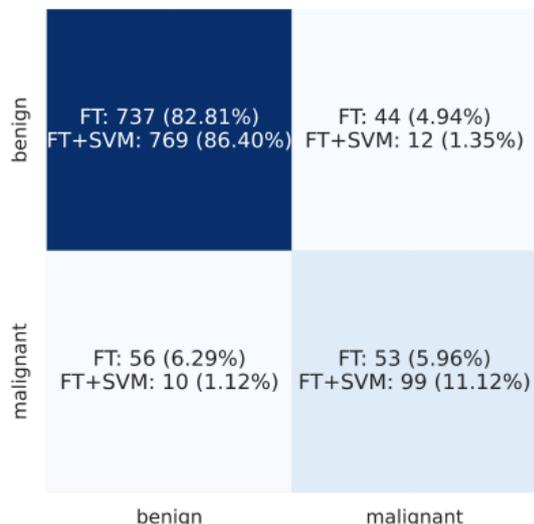


(b) InceptionV3 CM

Fig. 6. DDTI RoC curves: 15% of layers FT and FT+SVM

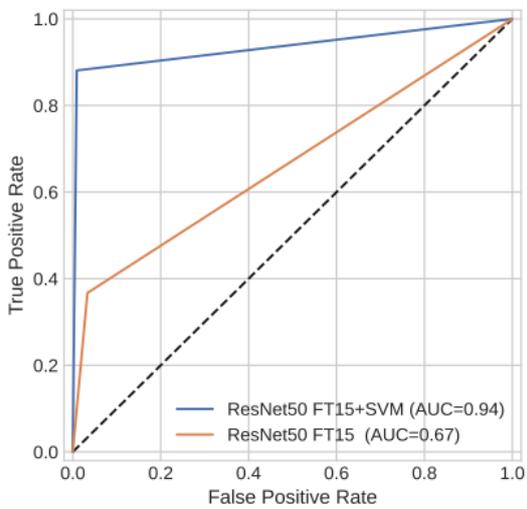


(a) ResNet50 CM

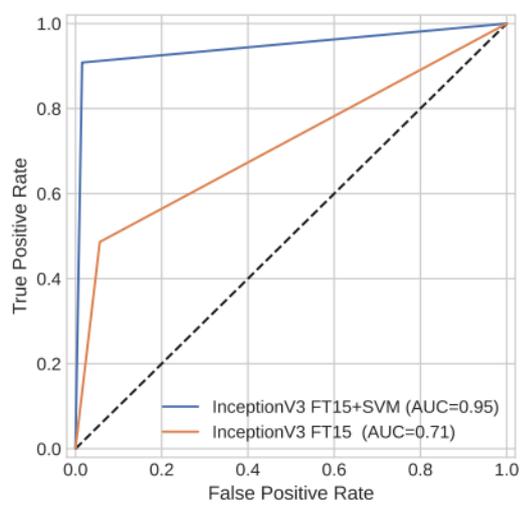


(b) InceptionV3 CM

Fig. 7. Salivary Glands confusion matrix: 15% of layers FT and FT+SVM



(a) ResNet50 RoC



(b) InceptionV3 RoC

Fig. 8. Salivary Glands RoC curves: 15% of layers FT and FT+SVM

4. Conclusion

This paper demonstrates the effectiveness of transfer learning and the effect of fine-tuning specific layers in enhancing ultrasound image classification tasks. Using different state-of-the-art ImageNet pre-trained models, various fine-tuning strategies were implemented and investigated such as fine-tuning 0%, 15%, 25%, 50%, 75%, 85%, and 100% of inner layers. The evaluation of these strategies was performed in two phases, first, fine-tuning with a basic classifier, and second, replacing the classifier with a linear SVM. The whole evaluation process has been implemented with a 5-fold cross-validation mechanism ensuring robust model evaluation and selection. Among the five implemented models, two models namely, ResNet50, and InceptionV3 have shown good performances while fine-tuning 15% of their layers. Additionally, the overall performance of these models has increased significantly while adopting a hybrid approach by leveraging a linear SVM on the classifier part of the fine-tuned models. The results of this study underscore the importance of optimizing deep learning techniques in ultrasound image analysis. Additionally, they shed light on the significance of fine-tuning strategies and classifier selection in achieving accurate and reliable classification outcomes.

Conflict of interest

No author claims to have any conflicts of interest.

References

- [1] Agarap A. F.: An architecture combining convolutional neural network (cnn) and support vector machine (svm) for image classification. arXiv preprint: 1712.03541, 2017.
- [2] Al-Dhabyani W. et al.: Dataset of breast ultrasound images. Data Brief. 28, 2019, 104863.
- [3] Bal-Ghaoui M. et al.: U-net transfer learning backbones for lesions segmentation in breast ultrasound images. International Journal of Electrical and Computer Engineering (IJECE) 13, 2023, 5747, [http://doi.org/10.11591/ijece.v13i5.pp5747-5754].
- [4] Cortes C., Vapnik V.: Support-vector networks. Machine learning 20, 1995, 273–297.
- [5] Kim H. E. et al.: Transfer learning for medical image classification: A literature review. BMC medical imaging 22(1), 2022, 69.
- [6] Kora P. et al.: Transfer learning techniques for medical image analysis: A review. Biocybernetics and Biomedical Engineering 42(1), 2022, 79–107.
- [7] LeCun Y. et al.: Handwritten digit recognition with a back-propagation network. Advances in neural information processing systems 2, 1989.
- [8] Mukhlif A. A. et al.: An extensive review of state-of-the-art transfer learning techniques used in medical imaging: Open issues and challenges. Journal of Intelligent Systems 31(1), 2022, 1085–1111.
- [9] Nanni L., Ghidoni S., Brahnam S.: Deep features for training support vector machines. Journal of Imaging 7(9), 2021, 177.
- [10] Pedraza L. et al.: An open access thyroid ultrasound image database. 10th International symposium on medical information processing and analysis 9287, 2015, 188–193.
- [11] Rodrigues P. S.: Breast ultrasound image. Mendeley Data 1(10), 2017, 17632.
- [12] Ronneberger O., Fischer P., Brox T.: U-net: Convolutional networks for biomedical image segmentation. 18 International Conference Medical Image Computing and Computer-Assisted Intervention – MICCAI, Munich, 2015, 234–241.
- [13] Salivary Glands Ultrasound Cases. Website [https://www.ultrasoundcases.info/cases/head-and-neck/salivary-glands/] (accessed: April 15, 2023).
- [14] Samee N. A. et al.: Deep learning cascaded feature selection framework for breast cancer classification: Hybrid cnn with univariate-based approach. Mathematics 10(19), 2022, 3631.
- [15] Shung K. K.: Diagnostic ultrasound: Past, present, and future. Journal of Medical and Biological Engineering 31(6), 2011, 371–374.
- [16] Srivastava R., Kumar P.: A cnn-svm hybrid model for the classification of thyroid nodules in medical ultrasound images. International Journal of Grid and Utility Computing 13(6), 2022, 624–639.
- [17] Tang Y.: Deep learning using linear support vector machines. arXiv, preprint: 1306.0239, 2013.
- [18] Wang Y. et al.: A hybrid classification method of medical image based on deep learning. Research Square, preprint, 2021.
- [19] Wang Y. et al.: The diagnostic value of ultrasound-based deep learning in differentiating parotid gland tumors. Journal of Oncology, 2022.
- [20] Xia X. et al.: Deep learning for differentiating benign from malignant parotid lesions on mri images. Frontiers in Oncology 11, 2021, 632104.
- [21] Yu X. et al.: Transfer learning for medical images analyses: A survey. Neurocomputing 489, 2022, 230–254.

M.Sc. Eng. Mohamed Bal-Ghaoui

e-mail: mohamed_balghaoui@um5.ac.ma

Received his Engineer's degree in Biomedical Engineering at the University of Mohammed V, ENSAM Rabat. He is a Ph.D. candidate researcher in Artificial Intelligence applied to Biomedical engineering. His research interest includes AI applied to Ultrasound Images for the early detection of cancer. He is part of E2SN Team – Biomedical Engineering Research Laboratory at ENSAM-Rabat.



<http://orcid.org/0000-0002-2143-6458>

Prof. My Hachem El Yousfi Alaoui

e-mail: h.elyousfi@um5r.ac.ma

Professor of Biomedical Engineering at the University of Mohammed V, ENSAM-Rabat, he is a member of the research laboratory E2SN – Biomedical Engineering Research Laboratory at ENSAM-Rabat, Mohamed V University in Rabat, Morocco. Members of the E2SN research group. Prof. El Yousfi is current research work is focused on biomedical data processing, AI, IoT and the hardware implementation of associated circuits.



<http://orcid.org/0000-0003-4285-0540>

Prof. Abdelilah Jilbab

e-mail: a.jilbab@um5r.ac.ma

Professor of Electrical Engineering at ENSAM Rabat of the University of Mohammed V in Rabat. He acquired his Ph.D. in Computer and Telecommunication from Mohammed V Agdal University, Rabat, Morocco in February 2009. He has published in the field of image processing, sensor networks, and signal processing for Parkinson's disease. His current interest is in embedded systems and wireless sensor networks (WSN) applied to biomedical. Dr. Jilbab is a member of the research laboratory E2SN – Biomedical Engineering Research Laboratory at ENSAM-Rabat at Mohamed V University.



<http://orcid.org/0000-0002-1577-9040>

Prof. Abdemasser Bourouhou

e-mail: a.bourouhou@um5r.ac.ma

Professor of Electrical Engineering at ENSAM, Mohammed V University Rabat. He acquired his Ph.D. in Electronics and Signal Processing from Ibn Tofail University in Kenitra, Morocco in April 2008. A. Bourouhou is a member of the Electronic Systems Sensors and Nanobiotechnology (E2SN) team of the STIS research center of UM5 Rabat. His research interests include biosignal processing, wireless sensor networks (WSNs) for environmental protection, and medical monitoring.



<http://orcid.org/0000-0002-6150-5374>