Ultrasound versus Elastography in the Study of Thyroid Nodules

Tiago Rocha
uc2018263670@student.uc.pt

Jaime Santos
jaime@deec.uc.pt

Alexandra André
alexandra.andre@estescoimbra.pt

José Silvestre Silva
jose.silva@academiamilitar.pt

Department of Physics, University of Coimbra, Portugal

University of Coimbra, CEMMPRE, ARISE, Department of Electrical and Computer Engineering, Portugal

Coimbra Health School, Coimbra, Portugal

CINAMIL & Portuguese Military Academy, Portugal

LIBPhys-UC & LA-Real, University of Coimbra, Portugal

Abstract

Thyroid nodules, despite appearing as a discrete lesion, constitute a prevailing pathological occurrence within the global population. The timely detection and diagnosis can help preventing the pathology from growing, minimising more severe effects on the human body. In this study, supervised machine learning and deep learning techniques were implemented to analyse ultrasound and elastography medical images increasing and improving the effectiveness of thyroid nodule detection. The results achieved using deep learning were superior to those achieved using machine learning. Specifically, for machine learning it was obtained a F1-Score of 97.20%, for the ultrasound images and a F1-Score of 75.40% using elastography images. Deep learning methodologies reached a F1-Score of 98.85% for ultrasound images and 89.15% for elastography images.

Keywords: Thyroid, ultrasound, elastography, traditional classifiers, Deep Learning

1 Introduction

Ultrasound and elastography are two important medical imaging tools widely used in the assessment of thyroid nodules (figure 1). Ultrasound is a medical imaging technique that uses sound waves to produce images of the thyroid and its nodules with good resolution, providing important information such as size, shape, structure, and the presence of fluids [1]. Elastography is a more recent medical imaging technique based on highly developed software that makes it possible to determine and assess changes in the structural properties of tissues and, consequently, their stiffness, which is crucial in differentiating thyroid nodules [2]. Ultrasound and elastography are safe, non-invasive medical imaging techniques and their use in the study of thyroid nodules is essential for making an accurate and effective diagnosis [3].

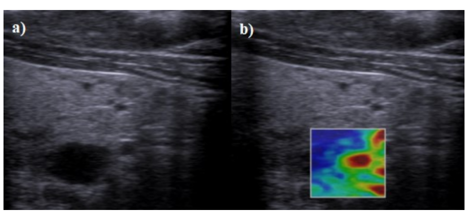

Figure 1 - Image of a thyroid nodule: a) Ultrasound; b) Elastography

2 Methods

2.1 Datasets

In total, 207 images were used, where 145 were collected directly by a health professional (76 ultrasound images and 69 elastography images with and without nodules) and 62 ultrasound images were retrieved from a public dataset, DDTI (Digital Database of Thyroid Ultrasound Images), which contains ultrasound images accompanied by a diagnosis made by radiologists and confirmed by biopsies when necessary, [4].

2.2 Classification and Feature Selection

After acquiring thyroid images of patients with and without pathology, the images are submitted to two classification models. In the case of machine learning, a sliding window is applied to the images to increase its number of images to make up for the shortage of images. This technique allows an image to be scrolled through a window with previously chosen dimensions from a starting point to the end of the image. It is also possible to define the step size, i.e., the number of pixels to move between each window. Then, the images are normalized and around 600 different features are extracted, which undergo a normalization process immediately before the best ones are chosen by the algorithm. The data achieved was processed by different classifiers such Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT) and K-nearest neighbors (KNN) to produce a variety of results to be compared. Cross-validation and classifier optimization mechanisms were also applied. The final aim is an implementation of classifiers to provide statistical results regarding the classification of thyroid images (with or without a nodule).

Regarding deep learning, to overcome the small dataset, the images were pre-processed, subjected to a data augmentation process and pre-trained on ImageNet (transfer learning). The classification model was implemented using a Convolutional Neural Network (CNN). The VGG16 is the neural network used in this study. This CNN has an architecture with 13 convolutional layers, 3 fully connected layers and 5 pooling layers. The first two types of layers mentioned above are associated with activation functions (ReLU), however the last layer is a SoftMax activation layer.

The metrics used to evaluate the results obtained by both above approaches were precision, accuracy, F1-Score and AUC (Area Under the ROC Curve).

3 Results and Discussion

A fixed number of features (20) was used to evaluate the performance of the different classifiers. Figure 2 illustrates the influence of step size and window size on the performance of the different metrics for the different classifiers.

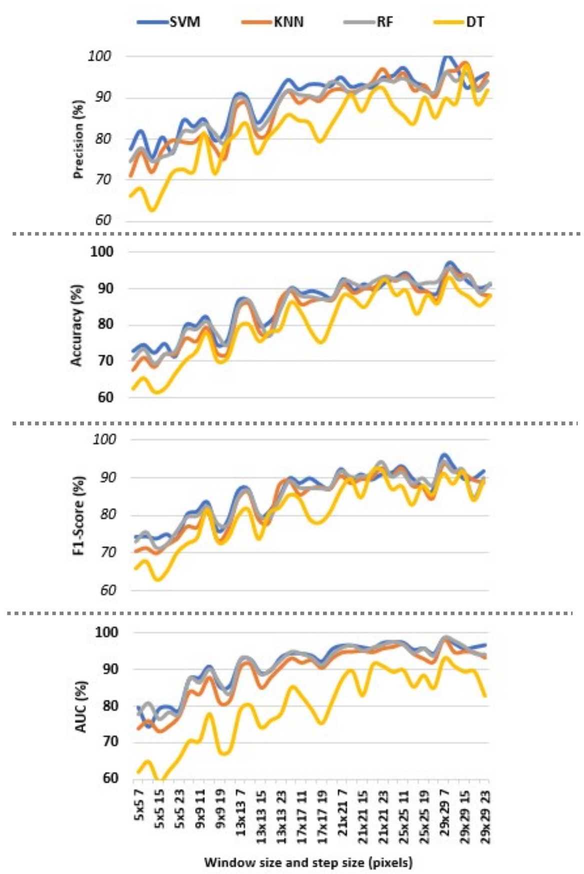

Figure 2 – Performance metric values by window size and step size for the different classifiers

It is observed that SVM classifier shows a better classification result overall, although in the AUC metric has a slight increase in performance on the part of the RF classifier. Therefore, the SVM classifier was used to acquire the subsequent results due to its good overall performance compared to the others.

The SVM was further tested to produce the maximum performance. The best values were achieved using a window of 25 x 25 (pixels) for a step size of 9. In the case of feature normalization, the StandardScaler and MinMaxScaler functions were studied and compared. The comparison included the incorporation of SelectKBest and Recursive Feature Elimination methods for feature selection, where the pair StandardScaler and Recursive Feature Elimination produced the best overall performance for the different metrics.

The performance of the classification model is closely linked to the number of features selected. Figure 3 shows the results of the classifier's performance in relation to variations in the number of features. In this case, only the F1-Score will be presented as a metric, to have a more detailed analysis of the ideal number of features.

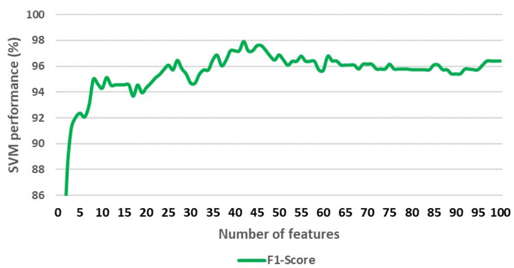

Figure 3 – SVM performance over the number of features

It is noticed a sudden increase in SVM performance as the number of selected features is 10. Then there is a less drastic rise with subsequent stabilization of performance. The SVM classifier performs best when the number of features is near 40. Finally, once all the variables had been explored and defined, the performance of the SVM was compared (figure 4) for the two groups of images (ultrasound and elastography).

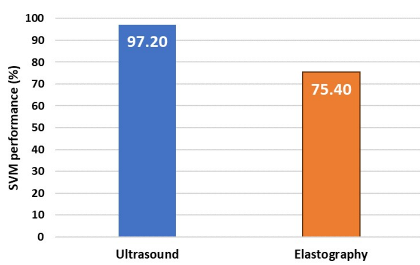

Figure 4 – SVM performance for ultrasound and elastography images

The images acquired through ultrasound produce a much better performance than the elastography. We also implemented a Deep Learning technique. After pre-processing the images, different parameters and configurations were tested to produce the best classification performance. In this context, a high number of epochs was selected to calculate an initial batch size. The analysis showed that a batch size of 16 yielded the best results. Using this established batch size, the ideal number of epochs needed was investigated, as shown in figure 5.

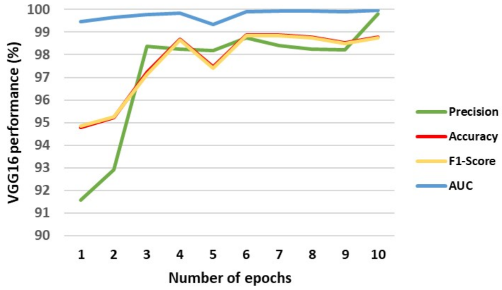

Figure 5 – VGG16 performance for different number of epochs

Examining the graph (figure 5), it becomes evident that the AUC metric always gives results that are too good, so it was promptly discarded. The performance of the accuracy metric increases, reaching its maximum at 10 epochs. The accuracy and F1-Score curves show a very similar behaviour, only varying in the order of tenths, with maximum performance peaks at 4, 6 and 8. Given the limitations of computational resources, 4 epochs were used in the following steps. As with machine learning, the metric used in the next tests was the F1-Score. A comparison between various optimisers was carried out. The results showed that the Adaptive Moment Estimation (Adam) optimiser performed better. Finally, the use of VGG16 in classifying thyroid nodules demonstrated superior performances, when complemented with a Multi-Head Attention method (figure 6).

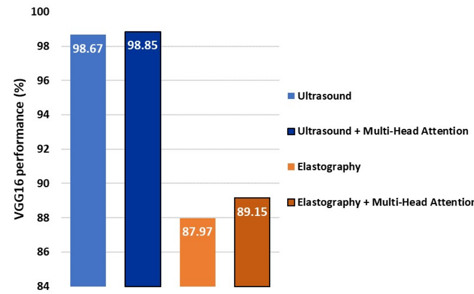

Figure 6 – VGG16 performance comparison

4 Conclusion

This work presents a comparative study of different classifiers as machine learning and deep learning, applied to the classification of ultrasound and elastography images of thyroid. The primary objective was to assess the performance of classification within this pathological context. The algorithms capacity to discriminate between images showing nodules and those devoid of such features was analysed. It was observed that machine learning exhibited a higher proficiency in classifying ultrasound-acquired images (97.20%) in contrast to those obtained through elastography (75.40%), with Support Vector Machine (SVM) emerging as the most suitable classifier. On the other hand, the use of deep learning, employing a Convolutional Neural Network (CNN) architecture, achieved a F1-Score of 98.85% for ultrasound images and 89.15% for elastography images. Machine Learning and Deep Learning stand as two pivotal techniques making easy the resolution of difficult problems characterized by substantial complexity and extensive data volume. The traditional classifier, while less intricate, needs limited data but requires manual programming for feature extraction. In contrast, Deep Learning operates on the premise of automated feature extraction processes. This distinction in feature acquisition processes potentially underlies the observed performance discrepancies in relation to elastography images. It is possible that the essential features required for accurate classification of elastography images using SVM might not be effectively extracted through the actual mechanism. To improve the results for elastography images a larger database should be used.

Acknowledgements

This research is sponsored by national funds through FCT – Fundação para a Ciência e a Tecnologia, under the project UIDB/00285/2020 and LA/P/0112/2020.

References

[1] C. Rumack and D. Levine, *Diagnostic ultrasound*, 5th ed. Philadelphia, USA: Elsevier, 2017.

[2] Barr Richard G., *Elastography: A Practical Approach*, 1st ed. New York, USA: Thieme, 2017.

[3] S. Tamhane and H. Gharib, "Thyroid nodule update on diagnosis and management," *Clin Diabetes Endocrinol*, vol. 2, no. 1, 2016, doi: 10.1186/s40842-016-0035-7.

[4] L. Pedraza, C. Vargas, F. Narváez, O. Durán, E. Muñoz, and E. Romero, "An open access thyroid ultrasound image database," in *10th International Symposium on Medical Information Processing and Analysis*, 2015. doi: 10.1117/12.2073532.