PULMONARY NODULES MALIGNANCY STRATIFICATION OVER COMPUTED TOMOGRAPHY (CT) SEQUENCES USING DEEP DISCRIMINATIVE REPRESENTATIONS

ALEJANDRA MORENO TARAZONA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2023

PULMONARY NODULES MALIGNANCY STRATIFICATION OVER COMPUTED TOMOGRAPHY (CT) SEQUENCES USING DEEP DISCRIMINATIVE REPRESENTATIONS

ALEJANDRA MORENO TARAZONA

Research work in partial fulfillment of the requirements for the degree of:
Magíster en Ingeniería de Sistemas e Informática

Advisor:
Fabio Martínez Carrillo
Ph.D in Systems and Computer Engineering

Co-Advisor:
Andrea Rueda Olarte, Ph.D
Ph.D in Systems and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2023

ACKNOWLEDGEMENTS

The author expresses her acknowledgment:

Mainly to God, and to my father Claudio Moreno, who has always supported me countless times to achieve all my goals, who is always there to give me a hand whenever I need it, who pushes me to fulfill my dreams, and who does not let me give up.

To my sister Kelly, for being my principal role model, who since I was a little girl has pushed me to follow the academy, to propose, to investigate, to learn, and who despite the distance shows up every day reminding me that things can be difficult, but there is always a reward in it. Also, to all my family, for always giving me their support and trust in me.

To my advisor, Professor Fabio Martínez, for having been by my side since my undergraduate, helping me with small questions and training me in this beautiful path of research, allowing me to question things, and to achieve what I thought at some point unattainable. From day one has believed in me and has believed that it is important to train scientists for this country. I will always be eternally grateful for the tremendous advisor role that he does.

To my co-advisor, Professor Andrea Rueda, who has helped me to increase my writing skills, the question formulation, and even gives me guidelines every time I felt lost, to form and grow as a researcher for the country.

To my research friends, especially Lina, Edgar, Santiago, and Andres, who has taught me to grow both personally and academically, with whom I started this research dream, and with whom I have gone through adversities and challenges, helping me to achieve this goal.

Final but not least, to the BIVL^{2}ab research group, the School of Systems and Computer Engineering, and the UIS, for allowing me to be in this alma mater and fulfill this dream.

Contents

List of Figures

pág.

List of Tables

LIST OF APPENDICES

pág.

ABSTRACT

TITLE: PULMONARY NODULES MALIGNANCY STRATIFICATION OVER COMPUTED TOMOGRAPHY (CT) SEQUENCES USING DEEP DISCRIMINATIVE REPRESENTATIONS [*]

AUTHOR: ALEJANDRA MORENO TARAZONA [**]

Keywords: DEEP DISCRIMINATIVE REPRESENTATIONS, CT, LUNG CANCER, LUNG NODULE CLASSIFICATION, ATTENTION MODULES.

Description: Lung cancer remains the principal cause of cancer-related deaths. Nodules are the main radiological finding, typically observed from low-dose CT scans. These masses are coarsely stratified according to textural and geometrical patterns, following criteria established by the Lung-RADS protocol. Nonetheless, the nodule characterization diagnosis remains subjective, reporting a moderate agreement among experts. Even worse, only the 5% of nodules have an associated correlation with lung cancer, which leads to a significant percentage of false positives during expert analysis. Today, computational approaches support diagnosis from nodule observations but principally between coarse malign and benign classes. Additionally, many nodules are unclassified due to their indeterminated nature. This work introduces a multi-attention architecture dedicated to multiple nodule classification, that takes advantage of local, intermediate, and non-local saliency maps, learned from independent branches. Validation includes an extensive analysis regarding multiple attention features, allowing to establish a correlation with other radiological findings. An agreement between four radiologists was also included in this study. The proposed approach achieves an AUC of 85.35% for a classical multi-classification and an AUC mean of 82.90% on a one-vs-all validation methodology. Both classification schemes were run into a k-fold cross-validation setup, showing competitive results in the state-of-the-art. After the training of the architecture was performed, the indeterminated nodules were mapped into the proposed method to support such challenging malignancy classification.

---

[*] DEGREE WORK

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Advisor: Fabio Martínez Carrillo, Ph.D. Co-Advisor: Andrea Rueda Olarte, Ph.D.

RESUMEN

TÍTULO ESTRATIFICACIÓN DE LA MALIGNIDAD DE LOS NÓDULOS PULMONARES EN SECUENCIAS DE TOMOGRAFÍA COMPUTARIZADA (TC) UTILIZANDO REPRESENTACIONES DISCRIMINATIVAS PROFUNDAS [*]

AUTOR: ALEJANDRA MORENO TARAZONA [**]

PALABRAS CLAVE: REPRESENTACIONES DISCRIMINATIVAS PROFUNDAS, CT, CÁNCER DE PULMÓN, CLASIFICACIÓN DE NÓDULOS PULMONARES, MÓDULOS DE ATENCIÓN

DESCRIPCIÓN: El cáncer de pulmón se mantiene como la principal causa de mortalidad relacionada con el cáncer a nivel mundial. Los nódulos son el principal hallazgo radiológico, típicamente observado a partir de tomografías computarizadas de baja dosis. Estas masas se estratifican de forma general según patrones texturales y geométricos, siguiendo los criterios establecidos por el protocolo Lung-RADS. No obstante, la caracterización del diagnóstico de los nódulos sigue siendo subjetivo, reportando un acuerdo moderado entre los expertos. Peor aún, únicamente el 5% de los nódulos tienen una correlación asociada con el cáncer de pulmón, lo que conduce a una alta tasa de falsos positivos durante el análisis de los expertos. Actualmente, los enfoques computacionales apoyan el diagnóstico a partir de las observaciones de los nódulos, pero fundamentalmente entre clases malignas y benignas. Incluso peor, muchos nódulos carecen de diagnóstico debido a su naturaleza indeterminada. Este trabajo introduce una arquitectura multi-atención dedicada a la clasificación múltiple de nódulos, que se beneficia de mapas de saliencia locales, intermedios y no locales, aprendidos a partir de ramas independientes. La validación incluye un extenso análisis respecto a las características de múltiple atención, permitiendo establecer una correlación con otros hallazgos radiológicos. Asimismo, se incluyó en este estudio una concordancia con múltiples radiólogos. El enfoque propuesto alcanza un AUC de $85,35\%$ para una multi-clasificación clásica y un AUC promedio de $82,90\%$ en un esquema de validación de uno contra todos. Ambos esquemas de clasificación se ejecutaron en una configuración de validación cruzada de k-fold, mostrando resultados competitivos en el estado del arte. Una vez realizado el entrenamiento de la arquitectura, los nódulos indeter-

---

[*] Trabajo de grado

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director: Fabio Martínez Carrillo, Ph.D.Co-Advisor: Andrea Rueda Olarte, Ph.D.

minados fueron mapeados al método propuesto para obtener un soporte en la clasificación de la malignidad.

INTRODUCTION

Lung cancer (LC) is the second most prevalent cancer worldwide ($\sim 11.4\%$), responsible for about 1.8 million deaths in the world and representing 18% of all cancer-related deaths[1]. Pulmonary nodules, small masses between $[3-30]$ millimeters, constitute the main biomarkers of LC, allowing the diagnosis, stratification, and follow-up of disease progression. These masses are typically observed in low-dose computed tomography (CT) scans, allowing to characterize each finding according to texture, morphology, and density[2]. To standardize LC screening, the American College of Radiology (ACR) has developed the Lung-RADS®[3] as a quality assurance tool used to support nodule grading. Nonetheless, the nodule analysis remains subjective, evidencing a dramatic low and moderate agreement among expert radiologists. For instance, a study with eight experts and 160 nodules report a mean κ coefficient of 0.51 (with a 95% confidence interval $[0.30-0.68]$) for inter-observer variability, and a mean κ of 0.57 (with a 95% confidence interval $[0.47-0.71]$) for intra-observer variability[4].

In fact, in clinical practice is very common to define nodules as indeterminated, or to have non-consensus among experts, delaying disease diagnosis. Although most indeterminate nodules could be benign, some of them could further turn out to be malign, with a ratio

---

[1] Hyuna Sung et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: vol. 71. 3. Wiley Online Library, 2021, pp. 209–249.

[2] Peter J Mazzone and Louis Lam. "Evaluating the patient with a pulmonary nodule: a review". In: *Jama* 327.3 (2022), pp. 264–273.

[3] Brady J McKee et al. "Performance of ACR Lung-RADS in a clinical CT lung screening program". In: *Journal of the American College of Radiology* 13.2 (2016), R25–R29.

[4] Sarah J van Riel et al. "Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management". In: *Radiology* 277.3 (2015), pp. 863–871.

between $5-60\%$[5]. Despite nodule size is correlated with the disease, a reported variability forces the use of the morphology, textural, and geometrical features to improve diagnosis and evaluation[6].

Several computerized vision-based methods have been proposed to classify pulmonary nodules as benign or malign, and in some cases, even trying to stratify their malignancy degree[7] [8]. A set of approaches are focused on exploiting the spatial information, allowing to extract textural nodule features from spatial slices[9] [10] or using 3D approaches to characterize the nodule appearance and geometry more concisely[11] [12]. Furthermore, attentional

---

[5] Pierre P Massion and Ronald C Walker. "Indeterminate pulmonary nodules: risk for having or for developing lung cancer?" In: *Cancer prevention research* 7.12 (2014), pp. 1173–1178.

[6] Rafael Paez, Michael N Kammer, and Pierre Massion. "Risk stratification of indeterminate pulmonary nodules". In: *Current Opinion in Pulmonary Medicine* 27.4 (2021), pp. 240–248.

[7] Shahin Heidarian et al. "Cae-transformer: Transformer-based model to predict invasiveness of lung adenocarcinoma subsolid nodules from non-thin section 3d ct scans". In: 2021.

[8] Ning Xiao et al. "Ensemble classification for predicting the malignancy level of pulmonary nodules on chest computed tomography images". In: vol. 20. 1. Spandidos Publications, 2020, pp. 401–408.

[9] Arnaud Arindra Adiyoso Setio et al. "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks". In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1160–1169.

[10] Mehdi Astaraki et al. "Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features". In: *Physica Medica* 83 (2021), pp. 146–153.

[11] Wentao Zhu et al. "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification". In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2018, pp. 673–681.

[12] Hong Liu et al. "Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification". In: *Journal of Digital Imaging* 33.5 (2020), pp. 1242–1256.

schemes have been codified to capture non-local morphology relationships[13] [14], and even attention-based techniques such as transformers have been used to locate and classify the nodule[15] [16]. These methods have achieved promising results to discriminate benign and malignant nodules, nevertheless, a large number of false positives may still arise; exposing the lack of robust and dense projections to characterize and quantify the nodules. Regarding multi-class nodule stratification, to bring a major sensitivity among cancer progression, some approaches have used hand-crafted features discriminated from classical machine learning strategies[17] [18] [19]. Deep representation has also introduced variational auto-encoders[20] or even an assembled convolutional neural network (CNN)[21]. The main challenge in such schemes is to define a specific class for indeterminated nodules. De-

---

[13] Hanliang Jiang et al. "Attentive and ensemble 3D dual path networks for pulmonary nodules classification". In: *Neurocomputing* 398 (2020), pp. 422–430.

[14] Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries".

[15] Jiancheng Yang et al. "Relational learning between multiple pulmonary nodules via deep set attention transformers". In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 1875–1878.

[16] Heidarian et al., "Cae-transformer: Transformer-based model to predict invasiveness of lung adenocarcinoma subsolid nodules from non-thin section 3d ct scans".

[17] Dmitriy Zinovev, Jacob Furst, and Daniela Raicu. "Building an ensemble of probabilistic classifiers for lung nodule interpretation". In: *2011 10th International Conference on Machine Learning and Applications and Workshops* 2 (2011), pp. 155–161.

[18] Matthew Yung et al. "Synthetic sampling for multi-class malignancy prediction". In: *arXiv preprint arXiv:1807.02608* (2018).

[19] Murillo B Rodrigues et al. "Health of things algorithms for malignancy level classification of lung nodules". In: *IEEE Access* 6 (2018), pp. 18592–18601.

[20] Devinder Kumar, Alexander Wong, and David A Clausi. "Lung nodule classification using deep features in CT images". In: *2015 12th conference on computer and robot vision*. IEEE. 2015, pp. 133–138.

[21] Xiao et al., "Ensemble classification for predicting the malignancy level of pulmonary nodules on chest computed tomography images".

spite some approaches including such criteria as an additional label, there is no clear support to aggregate uncertainty in clinical decisions.

To pave the way for proper stratification of pulmonary nodule malignancy, among different disease stages, a multi-volumetric-head multi-scale self-attention (called multi-attention) architecture is introduced in this work. The proposal integrates spatial representations through multi-scale projections to capture non-local regions, providing a proper nodule characterization and quantification based on texture, geometry, size, and morphology features. In addition, several attentions are classified in parallel, assisting in classification by obtaining more precise non-local relationships, which can be further correlated with radiological features. The main contributions can be summarized in the following aspects:

1. A multi-attention architecture based on a multi-scale decomposition, achieving intermediate and non-local relationships that enhance the feature maps with respect to texture, morphology, and size.

2. An exhaustive examination of radiologists' agreement levels considering the malignancy diagnosis.

3. To support the indeterminate nodule classification, such observations were projected on the multi-attention architecture to retrieve a malignancy degree.

1. FUNDAMENTALS AND PREVIOUS WORK

1.1. Lung cancer

There are two sub-types of lung cancer, small cell (SCLC) and non-small cell (NSCLC). The survival rate for SCLC is around [14%-15%], while NSCLC is the most frequent type of cancer, and its average mortality rate is around 58%[22]. On the one hand, SCLC cancer (with findings on a micron-scale) is directly analyzed from a biopsy. On the other hand, for NSCLC cancer, the nodules (that range from 3 to 30 mm) are the main biomarker of the disease, and their respective analysis is carried out from histopathological and radiological standards.

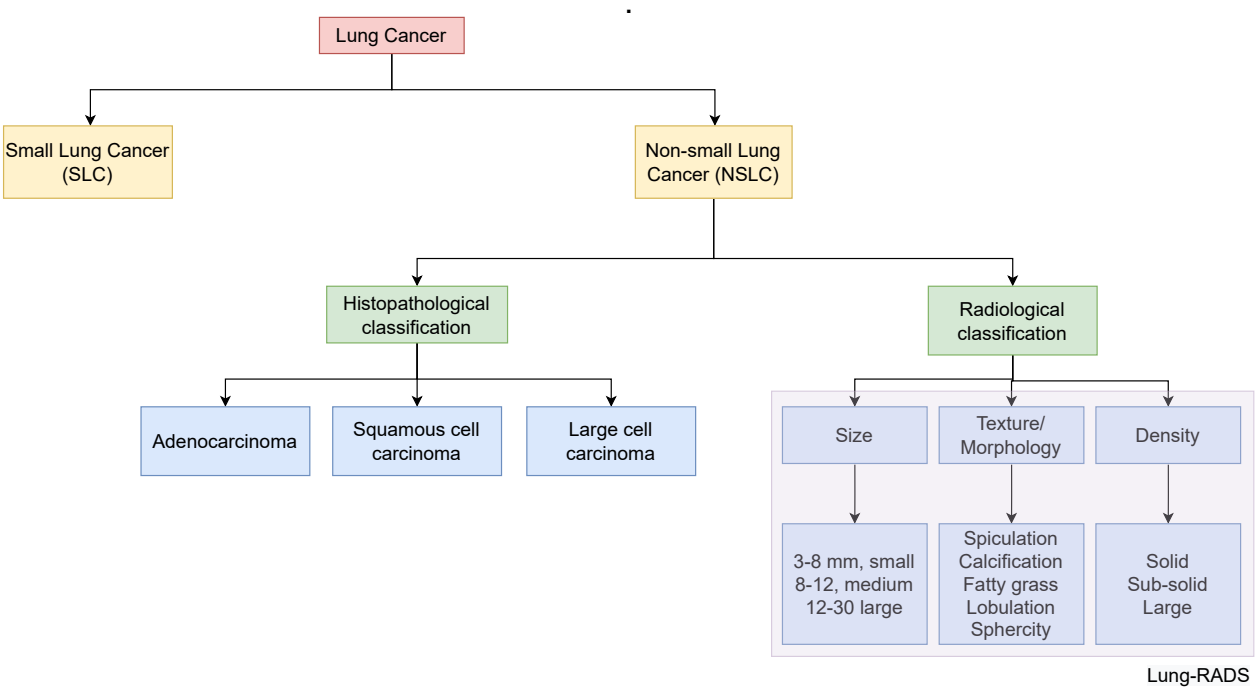
Figure 1 illustrates the two analyses of such non-small nodules and the considered primitives in each standard. From the histopathological analysis, the nodules are categorized as carcinoma, adenocarcinoma, and squamous features. From radiological observations, the Lung-RADS® protocol defines a set of geometrical and morphological primitives that allows to characterize and determine the degree of cancer of each nodule. It should be noted, that for some clinical protocols, isolated radiological features, such as the size may be used as an index of malignancy of the disease. For instance, nodules larger than 20 mm have a suspicion of malignancy of 80%, while nodules measuring ten mm have a probability of malignancy of 15.2%[23]. In the next section, we detail histological and radiological standard protocols to analyze nodules.

---

[22] Rebecca L. Siegel et al. “Cancer statistics, 2020”. In: *CA: A Cancer Journal for Clinicians* 70.1 (2020), pp. 7–30. DOI: 10.3322/caac.21590. eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21590. URL: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21590.

[23] Ross H Albert and John J Russell. “Evaluation of the solitary pulmonary nodule”. In: *American family physician* 80.8 (2009), pp. 827–831.

Figure 1. Stratification of nodules depending on histological or radiological classification.



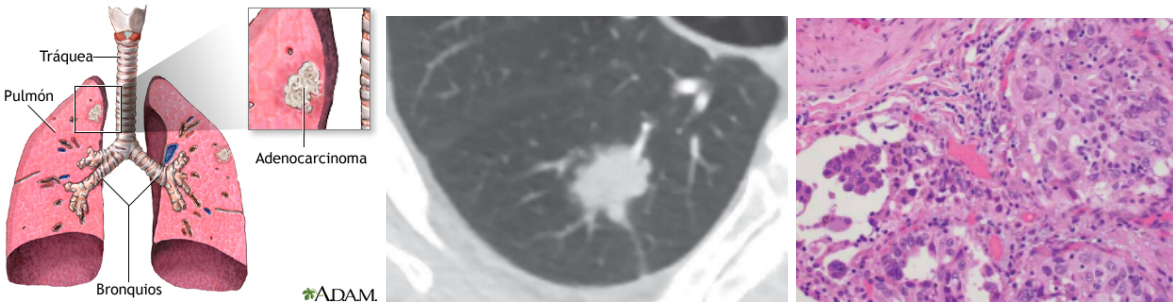
1.1.1. Histopathological classification From microscopical observations, the NSCLC lung cancer has three main sub-types of lung cancer (LC): adenocarcinoma, which accounts for 40% overall for all existing sub-types of LC (seen in Figure 2); the squamous cell carcinoma with 25% of all the LC sub-types (Figure 3); and the large cell carcinoma with the lowest percentage of LC (10%) (Figure 4). There exist other types but appear with much less frequency[24]. An explanation for each of the three main sub-types is described as follows:

- **Adenocarcinoma:** namely appear from tumors located peripherally on lungs, which report detectable mucin production by secreting digestive juices, mucus, and other fluid

[24] Jaime Rodriguez-Canales, Edwin Parra-Cuentas, and Ignacio I Wistuba. "Diagnosis and molecular classification of lung cancer". In: *Lung Cancer* (2016), pp. 25–46.

types. This cancer is generally found in a patient with a smoke background. From histopathological observation, it forms in glandular epithelial cells and has different histologic growth patterns, such as solid as the most common, micropapillary, papillary, acinar, and lepidic. Figure 2 illustrates a typical radiological and histological observation of this cancer type.

Figure 2. Adenocarcinoma example[25]. a) Lung images localization, b) a CT finding observation, c) the corresponding histological adenocarcinoma pattern.



a) b) c)

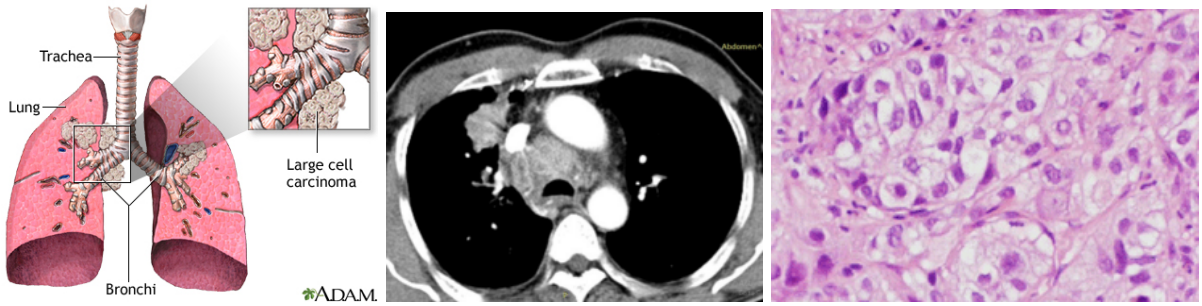
- **Squamous cell carcinoma:** is usually located at the center of the parenchyma concerning the main or lobar bronchus. Histologically, this tumor presents intercellular bridges and has nests of squamous epithelial cells emerging from the epidermis and spreading into the dermis, and it is also common in smokers. Figure 3 illustrates the localization, finding observations from CT, and the corresponding histological image.

- **Large cell adenocarcinoma:** is also defined as undifferentiated non-small cell carcinoma since it shows no evidence of squamous cell or glandular differentiation. Histologically it presents sheets and nests from moderate to abundant amounts of the cytoplasm of large cells. Figure 4 is shown the typical localization, the corresponding CT finding, and the respective histopathological pattern.

Figure 3. Illustrates squamous cell carcinoma images[26]. a) typical localization of this cancer type b) an example of CT finding, and c) the corresponding histology.

a) b) c)

Figure 4. A Large cell carcinoma illustration[27]. In a) is shown the typical localization, in (b) the corresponding CT finding associated with this cancer, and in (c) the corresponding large cell carcinoma histology.



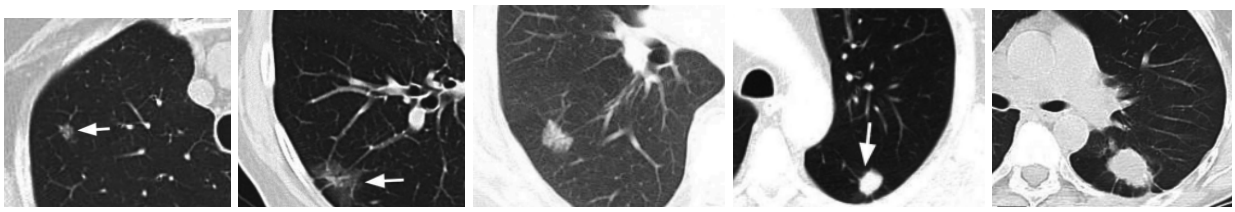
a) b) c)

1.1.2. Radiological classification The nodules can be identified through various visual and/or morphometric characteristics such as internal structure, spiculation, texture, lobulation, and sphericity, among others[28]. In fact, the Lung-RADS protocol group these findings and provide a likelihood to stratify nodule malignancy. Particularly, this protocol is based on the size, morphology, and density of nodules to provide an alphanumeric score ranging

[28] Annette McWilliams et al. "Probability of cancer in pulmonary nodules detected on first screening CT". in: *New England Journal of Medicine* 369.10 (2013), pp. 910–919.

from 0 to 4 as shown in Figure 5. Table 1 is summarized the findings associated with each Lung-RADS score-level.

Figure 5. Differentiation of all lung-RADS stages. All the images were taken from Figueroa *et. al*[29]. (a)Lung-RADS stage 2, (b)Lung-RADS stage 3, (c)Lung-RADS stage 4, (d)Lung-RADS stage 4A, and (e)Lung-RADS stage 4X.



a) b) c) d) e)

Figure 6 is illustrated some observational CT examples at different Lung-RADS stages. In such examples, they can be observed the different sizes of such nodules and the textural differentiation that provide observational tools to the radiologist to categorize the nodules. Regarding nodule features, Figure 6 is provided some examples of different nodules that expose diverse morphologies that support malignancy classification. For instance, the nodules in Figures 6-a,b have lobulated features with smooth borders and some calcification patterns. In Figure 6-c,d are presented some spiculated nodules with an irregular shape that have a higher malignant probability. Contrary, the nodules in figure 6-e,f have benign patterns but may correspond to false positives. The Lung-RADS protocol is an effort to reduce subjectivity in the diagnosis and to generalize the analysis of different nodule findings. Nonetheless, there exists reported a significant number of false-positive cases, and a miss correlation with biopsy results. This fact is associated with the relative low sensitivity of the scale and the coarse classification of nodules. Therefore, computational aid support is demanded to help with the characterization of nodules and to measure the progression of such regions[30].

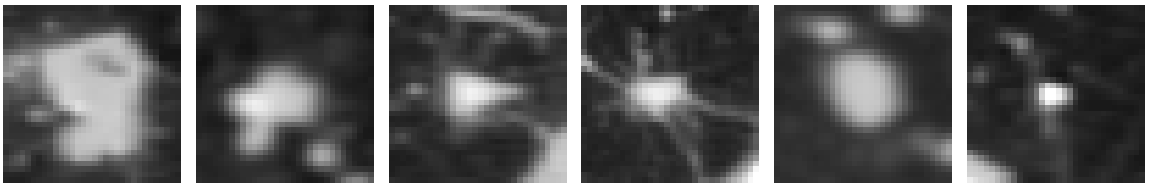
[30] Paul F Pinsky et al. “Performance of Lung-RADS in the National Lung Screening Trial: a retrospective

Table 1. The scores proposed by the Lung-RADS protocol, depend on the shape and/or size of the pulmonary nodule.

Lung-RADS Score	Meaning
LR-0 or incomplete	This score is associated with studies that are inadequate or incomplete and therefore a risk value cannot be assigned.
LR-1 or negative	This score is associated with studies without nodule findings or with benign nodules (composed of calcification of fatty grass). The probability of developing cancer remains at 1%. A further CT study is recommended to follow nodule progression.
LR-2 or benign appearance	This score is associated also with the patient with benign nodules (1% chance to evolve cancer) but it is recommended a CT study in one year.
LR-3 or probably malign	This score is assigned to a patient with suspicious nodules (between 6 to 8 mm) that requires a six-month following. In this stage, there is a 5% of chance developing cancer.
LR-4 or suspicious	At this level, it is included studies with large nodules that may be growing progressively. These nodules have spiculations, and asymmetric interlobular septal thickening, among other characteristics. The probability to develop cancer at this stage is around 37%.
	LR-4A: At this level are reported nodules between 8-15 millimeters, being the solid ones larger than six millimeters. These nodules should be followed exhaustively. The probability of developing cancer is between 5-15%.
	LR-4B: The patients at this level have a chance greater than 15% of developing cancer in the following year. This level is associated with nodules larger than 15 millimeters if they are solid, and larger than 8 millimeters if they are part-solid.
	LR-4X: This stage-score is referred to thoracic images that are suspicious but have not yet been specified, examples of which may be pleural effusion, spiculated nodules, thick-walled cystic lesions, interlobular septal thickening, among others.

assessment". In: *Annals of internal medicine* 162.7 (2015), pp. 485–491.

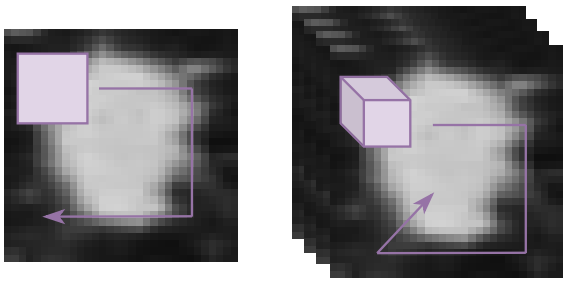
Figure 6. Radiomic characteristics of lung nodules. (a),(b) represents a lobulated nodule, (c),(d) a spiculated nodule, and (e),(f) smooth nodules



1.2. Deep representations

1.2.1. Convolutional-based methods Convolutional neural networks (CNNs) are models for learning classification tasks through generative and descriptive approaches. Normally, a CNN consists of an input layer, an output layer, and the so-called hidden layers. These hidden layers are namely convolutional layers, allowing to obtain a dense and robust representation of the classification task. In addition to this, after these convolutional layers, a max or average pooling operation is generally used, activating those important values of the image, allowing to obtain a better learning representation.

Figure 7. Convolutions depending on the kernel size. A) A normal spatial convolution is denoted as 2D convolution. B) A spatio-temporal convolution, which integrates each slice of the pulmonary nodules to obtain a wider spectrum of the morphology denoted as 3D convolution.



a) 2D Convolution b) 3D Convolution

In fact, CNNs can obtain a different categorization depending on the kernel dimensionality applied to the image, being spatial (2D) or volumetric (3D) as shown in Figure 7. The 2D CNNs allow to obtain the spatial information of the image, capturing local and regional information through the nodule, providing different relationships around the morphology, which are mostly used in classification tasks[31] [32]. On the other hand, 3D CNNs are specialized in obtaining volumetric information through medical scans. In the nodules context, it is widely used due to the complete morphology on the pulmonary nodule (x, y, z). In this case, (x, y) represent the nodule size in one slice, meanwhile, z has the whole information of each slice in order to reconstruct all the nodule, in order to find several, and more robust relationship patterns, taking into account the size and edges of the nodule structure[33] [34].

1.2.2. Attention-based methods Convolutional representations focus on computing local dependencies, which results in attractive in the image domain to avoid, for instance, translation and rotation dependencies. This inductive bias in convolutional models implies locality during the modeling. Nonetheless, these convolutional representations remain limited to capturing non-local dependencies that may enrich modeling key relationships and solve particular problems. For instance, in the study of a particular shape, recovered from an image observation, it would be interesting to associate relationships between non-local boundaries to recover morphological patterns that include such dependencies. Complementary, the attention-based models were proposed for natural language process-

---

[31] Setio et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks".

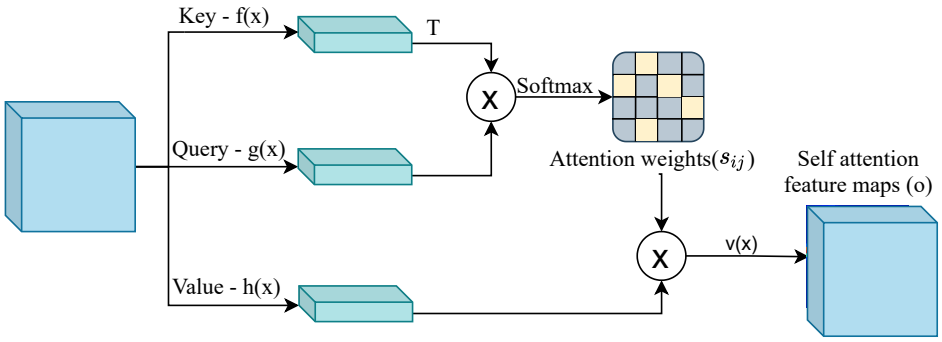
[32] Astaraki et al., "Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features".

[33] Qi Dou et al. "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection". In: *IEEE Transactions on Biomedical Engineering* 64.7 (2016), pp. 1558–1567.

[34] Zhu et al., "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification".

ing (NLP) schemes, exploiting representation that captures non-local relationships over sequence-to-sequence problems[35] [36]. Indeed, attention is derived from the weighting of the generated feature maps, according to keys that provide guidance in localization, characterization, and quantification. Recently, attention modules have focused on improving the CNN's performance in different ways, taking into account both image spatiality[37] [38] and image non-locality[39], trying to compute diverse types of correlations between pixels, which helps to improve the classification task.

Figure 8. Methodology proposed by[40] showing the convolutional self-attention performance.



As can be seen in Figure 8, the self-attention generally consists of several non-linear pro-

---

[35] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. "Effective approaches to attention-based neural machine translation". In: *arXiv preprint arXiv:1508.04025* (2015).

[36] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate". In: 2014.

[37] Long Chen et al. "Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5659–5667.

[38] Sanghyun Woo et al. "Cbam: Convolutional block attention module". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

[39] Han Zhang et al. "Self-attention generative adversarial networks". In: *International conference on machine learning*. PMLR. 2019, pp. 7354–7363.

jections called the key ($f(x) = W_k x_i$), the query ($g(x) = W_g x_i$) and the value ($h(x) = W_h x_i$) which project the input feature map (x) through the weights W_f, W_g and W_h respectively. From the key and the query is possible to calculate the attention matrix that encodes the non-local correlation between $f(x)$ and $g(x)$ across all the features per pixel of the nodule. These projections can be 2D convolutions or 1×1 convolutions. Such matrix is then computed through a similarity operation (s_{ij}) called "alignment score" across two pixels of the descriptor (x_i, x_j). Thus s_{ij} could be a linear operation, bi-linear operation, or even a multi-layer perceptron. After the s_{ij} is calculated, a softmax is applied to the matrix which represents the highest weighting values obtaining an "attention weights". The $h(x)$ can also be modeled as a 2D convolution or even a 1×1 convolution. In this case, a calculation of the value projection will be performed concerning the attention weights, to weight each activation map obtained in s_{ij} to obtain the self-attention map needed (o).

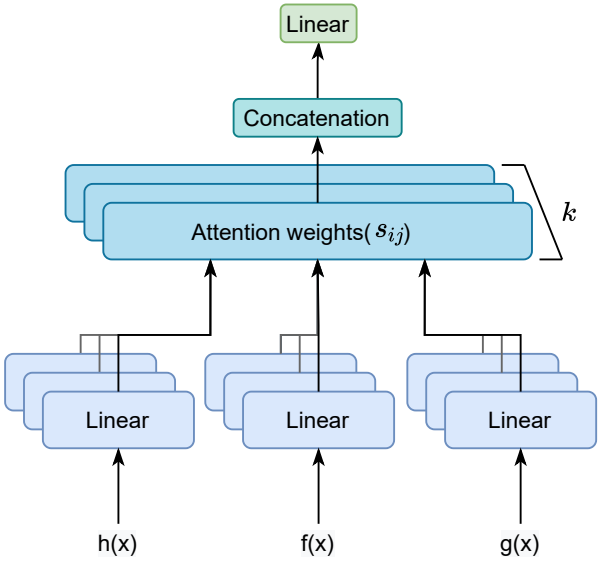
1.2.3. Multi-head attention methods An image can provide multiple relationships, which can be related to a specific class. This cannot be easily achieved only with a self-attention feature map, therefore, by computing different attention outputs in parallel, it will be possible to obtain several projections and multiple relationships that allow a better classification.

The following k-dimentional attention weights s_{ij}^k are calculated in this case, as can be seen in Figure 9 through different projections (k) carried out by $f(x), g(x),$ and $h(x)$ of dimensions k_f, k_g, and k_h respectively obtaining multiple non-local relationships. These projections will then be multiplied by the value in order to produce the attention matrix which is formally defined as:

$$\text{Attention}(g, f, h) = \text{softmax}\left(\frac{gf^T}{\sqrt{k_f}}\right) h \tag{1}$$

As mentioned above, given that the calculation of the k-dimentional $f(x)$, $g(x)$, and $h(x)$, k attention matrices are then obtained, capturing even more relevant or important informa-

Figure 9. Multi-head attention, with k projections and attention weights in order to obtain a multi-relationship attention map.



tion that allows weighting the feature maps. Therefore, a $multi\text{-}head(g, f, h)$ is obtained, which have multiple concatenations of diverse attentions called "head" concatenated in the form: $concat(head_1, head_2, ..., head_k)$ achieving to have a higher correlation between pixels[41].

1.3. Lung nodule characterization from computational approaches

LC diagnosis has been supported by different computational strategies that characterize and automatically stratify nodule masses, observed from CT studies. A first group of approaches has been dedicated to binary discriminate malignancy from 2D spatial nodule features. For instance, implementing a multi-view convolutional architecture to enrich nodule representation from nine patches in planes corresponding to the cube symmetry

[41] Ashish Vaswani et al. "Attention is all you need". In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.

plane[42]. Likewise, Causey *et al.* proposes a spatial convolutional network by varying the sizes of the architectures to provide more context in the classification task[43]. Along the same line, the use of dilated convolutional blocks allowed to recovers regional context relationships, exploiting textural correlations at different field scales[44]. Lastly, other strategies have supported nodule characterization by including a dual path net that includes the image context information but also the nodule segmentation to deal with intra-nodule features[45]. The explicit shape information deals with false positive reduction but may result unrealistic in clinical scenarios to delineate each nodule mass for further analysis.

Following the same binary discrimination, other strategies take advantage of volumetric nodule structures from 3D convolutional representations, integrating dual-paths with nodule segmentation's[46]. Likewise, Dey *et al.* introduced a volumetric convolutional scheme that follows a DenseNet backbone adapted to receive a dual-path input with the nodule cropped by using regions of interest at two different spatial sizes[47]. From such multiple inputs, the approach covers the global and local nodule context. Such net also recovers partial nodule predictions, adapting output layers at different intermediate layers, which

---

[42] Setio et al., "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks".

[43] Jason L Causey et al. "Highly accurate model for prediction of lung nodule malignancy with CT scans". In: *Scientific reports* 8.1 (2018), pp. 1–12.

[44] Mundher Al-Shabi, Hwee Kuan Lee, and Maxine Tan. "Gated-dilated networks for lung nodule classification in CT scans". In: *IEEE Access* 7 (2019), pp. 178827–178838.

[45] Astaraki et al., "Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features".

[46] Zhu et al., "Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification".

[47] Raunak Dey, Zhongjie Lu, and Yi Hong. "Diagnostic classification of lung nodules using 3D neural networks". In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE. 2018, pp. 774–778.

further are integrated to establish a final prediction. Similarly, an ensemble approach was adapted with a total of 3D convolutional nets to bring a nodule malignancy prediction, which in turn uses multiple nodule observations to capture nodule context[48]. Such multiple input scale observations may include surrounding features that could bias the representation to neighboring regions.

Recently, some approaches have designed attention methodologies to enrich nodule description from non-local relationships, aiming to cover a large spectrum of radiological finding features. For instance, some proposals include the convolutional block attention module (CBAM)[49] to integrate channel attention features and summarize convolutional responses, which are thereafter modeled with spatial information to recover standing-out nodule patterns[50] [51]. Also, Fu *et al.* introduced an approach that selects multiple slices from each nodule, according to radiometric features, to compute a specific content of attentions, which are further integrated into cross-attributed attentions for multiple morphological predictions[52]. In fact, self-attention mechanisms have been also implemented to achieve a non-local nodule representation from 3D convolutional backbones, which in turn integrate additional processing paths to include spatial attention at different scales[53]. In a more enriched representation, the transformer architectures have been adapted for nod-

---

48 Liu et al., "Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification".

49 Woo et al., "Cbam: Convolutional block attention module".

50 Sung et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries".

51 Hanliang Jiang et al. "Learning efficient, explainable and discriminative representations for pulmonary nodules classification". In: *Pattern Recognition* 113 (2021), p. 107825.

52 Xiaohang Fu et al. "Attention-Enhanced Cross-Task Network for Analysing Multiple Attributes of Lung Nodules in CT". in: 2021.

53 Jiang et al., "Attentive and ensemble 3D dual path networks for pulmonary nodules classification".

ule modeling, using multiple attention mechanisms at the same level of processing (known also as multi-head)[54] [55]. These works report a remarkable capability to discriminate malignant nodules from benign masses in a binary classification problem. Nonetheless, there is no evidence of its capability to discriminate among nodule stages. In fact, nodule stratification results as a key to support clinical decisions in scenarios with high uncertainty in the determination of the disease level.

The classification among multiple disease stages, according to Lung-RADS® protocol, has been addressed from some computational strategies, following classical machine learning algorithms and handcrafted features from nodule observations[56] [57]. These approaches are however limited to the particular assumptions of designed descriptors to recover textural or geometrical properties of the observations. Also, deep auto-encoder representations have been proposed in the literature to design continuous and progressive metrics to categorize nodule malignancy[58]. Some approaches have coarsely grouped malignant and benign classes to solve a binary problem, following the fusion of deep features and handcrafted descriptors[59]. In a similar way, Xiao *et al.* proposes a late fusion model using a denoising auto-encoder (ResNet-18) and handcrafted features to classify among

---

[54] Yang et al., "Relational learning between multiple pulmonary nodules via deep set attention transformers".

[55] Dongxu Liu et al. "Res-trans networks for lung nodule classification". In: *International Journal of Computer Assisted Radiology and Surgery* (2022), pp. 1–10.

[56] Yung et al., "Synthetic sampling for multi-class malignancy prediction".

[57] Rodrigues et al., "Health of things algorithms for malignancy level classification of lung nodules".

[58] Kumar, Wong, and Clausi, "Lung nodule classification using deep features in CT images".

[59] Yutong Xie et al. "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT". in: *Information Fusion* 42 (2018), pp. 102–110.

five nodule severity classes[60]. In such approaches, a strong assumption is followed by aggregating the indeterminate nodules to benign and malignant classes, in independent experiments, to enrich the training setup.

Typically, the radiologist uses five classes to stratify nodule malignancy (highly benign, moderately benign, indeterminated, moderately malign, and highly malign). However, a large percentage of the analyzed nodules are labeled as indeterminated (without a clear degree association) due to a mixture of nodule findings. For most of the computational approaches, including such indeterminated classes in the modeling results adds a large bias in the classification and limits the supporting information relevant to the clinical routine. As an approach, a collaborative deep learning scheme has been introduced[61], by integrating multiple nodule views which are mapped to a set of UNet architectures to recover nodule textural properties. Such nodule features are then integrated with multiple convolutions to obtain a binary classification. The multi-classification is then approximated by following a one-versus-all scheme. In the same way, Roy *et al.* implements multiple UNets, enriched with self-attention mechanisms, to generate nodule descriptors; which thereafter are discriminated from a convolutional net and adjusted with multiple classification tasks (real vs fake, benign vs malignant, and nodule severity)[62]. In validation, nodule severity estimation also follows a one-versus-all scheme. These strategies could require the estimation of a huge amount of parameters, which may over-fit the representation and collapse to specific data inputs.

---

[60] Xiao et al., "Ensemble classification for predicting the malignancy level of pulmonary nodules on chest computed tomography images".

[61] Yutong Xie et al. "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT". in: *IEEE transactions on medical imaging* 38.4 (2018), pp. 991–1004.

[62] Rukhmini Roy, Suparna Mazumdar, and Ananda S Chowdhury. "ADGAN: Attribute-Driven Generative Adversarial Network for Synthesis and Multiclass Classification of Pulmonary Nodules". In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).

2. RESEARCH PROBLEM

Pulmonary nodules are the main biomarkers of lung cancer, being classified as malignant depending on the type of density and morphometric characteristics. Low-dose CT scans nowadays provide a complete visualization of the morphology of the pulmonary nodule, being the principal tool to carry out a diagnosis of the disease. However, remarkable intra- and inter-observer variabilities in nodule classification by CT scans have been reported. In fact, studies find a κ of 0.51 and 0.57 for intra-observer and inter-observer variability, respectively. Furthermore, there is a high probability that pulmonary nodules around three to eight millimeters can present as false positives. A dramatic point in such misclassification is that potential malign nodules are not sent to histopathological analysis, losing an early and effective treatment in patients.

Several computational methods have emerged to support the pulmonary nodule classification. Multiple approaches, such as spatial and temporal convolutional networks, have been proposed to pursue local relationships. These representations are dependent on locality modeling, which many times may be prohibitive to enhance nodule shape properties. More recently attention modules have been achieved to the non-local characterization of nodules, showing remarkable results w.r.t convolutional representations. These current architectures nonetheless are insufficient to recover the geometry signature of nodules that allows to discriminate among stages of the disease. Also, major studies should be carried out to study major learning contributions in such feature maps, regarding the radiological findings described in standards and protocols.

Research question: How can a discriminative model be developed to characterize nodular patterns according to the degree of nodule malignancy?

3. OBJECTIVES

General Objective

- To propose a deep discriminative representation to characterize lung nodules, according to their degree of malignancy, on computed tomography (CT) scans.

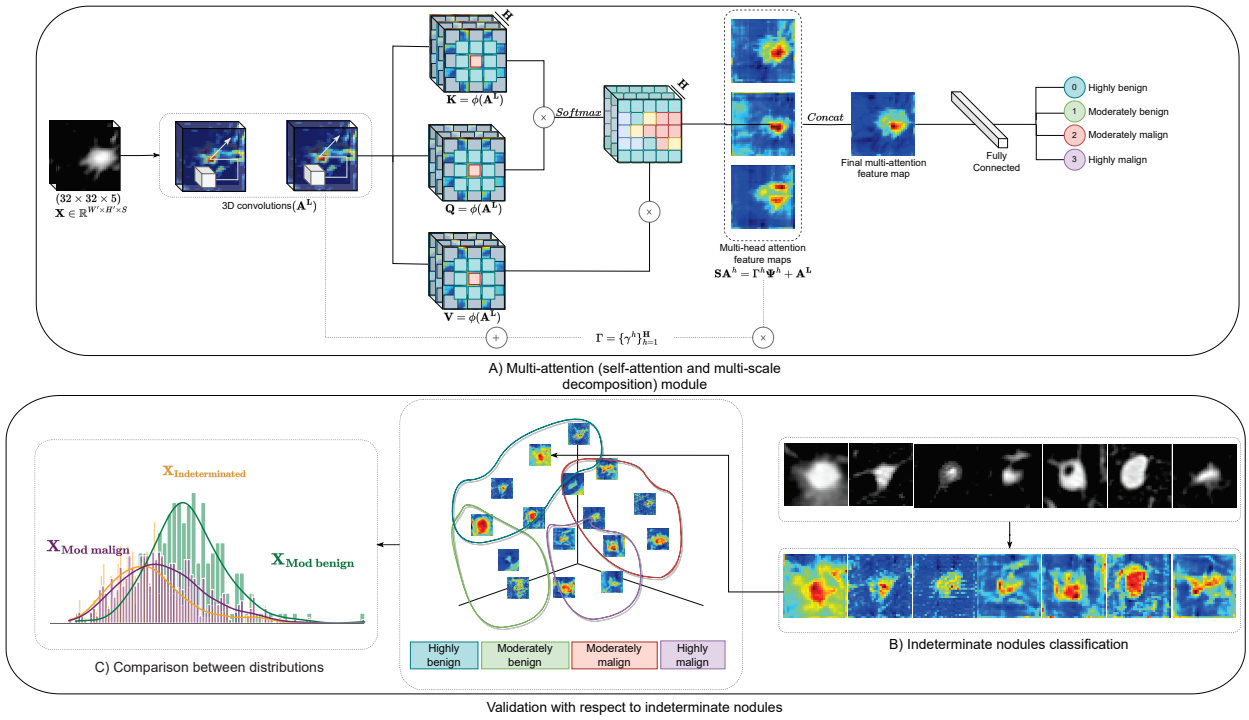
Specific Objectives

- To select an anonymized CT-scan dataset that has information related to nodule localization and categorization according to malignancy degree.
- To develop a discriminative deep model to classify cropped lung nodules observations, according to the degree of malignancy.
- To integrate an attentional nodule scheme to retrieve lung nodule saliency maps and obtain non-local relationships of nodule shapes.
- To validate the stratified discriminative model regarding the capability to support classification of benign and malign nodules.

4. PROPOSED APPROACH

This work introduces an architecture with multiple attention mechanisms to robustly characterize non-local volumetric nodule observations and deal with the classification of four severity levels of the disease. From the resultant embedding representation, indeterminated nodules are associated with the most probable class, as a mechanism to support clinical decisions. The pipeline for the proposed approach is summarized in Figure 10.

Figure 10. Pipeline of the proposed method. A) The multi-attention (multi-head and multi-scale) module. B) After training, a set of indeterminate nodules are classified, together with an exploration of morphologies. C) Statistical comparisons between actual malign or benign distributions and inferred classes are finally performed.



4.1. A 3D convolutional representation

Nodules are typically observed across CT axial slices and even supported from complementary views due to the volumetric morphology of such masses. From such considerations, we then assume a volumetric nodule representation captured from 3D regions of interest (RoIs) that bound abnormal masses. Here, the RoI input is then defined as the volumetric region $\mathbf{X} \in \mathbb{R}^{W' \times H' \times S'}$, with spatial dimensions ($W' \times H'$) and S' slices. The effort to process such 3D regions lies in the capability to recover robust shape and textural representation, projected at different axis directions, that together constitute a descriptor associated with different levels of the disease.

Such volumetric inputs are then projected into a convolutional backbone, conformed by several 3D convolutional layers hierarchically organized, that learn and enrich local nodule representation. Specifically, the RoI $\mathbf{X}$ achieves a hierarchical decomposition of visual information, recovered as a set of activation maps $\mathbf{A^1}, \ldots, \mathbf{A^L}$, organized at different layers $\mathbf{L}$. In such case, $\mathbf{A^{L_i}} = a^1, a^2, ..., a^J$, is formed by a set of J convolutional feature maps that code textural patterns (see in Figure 10). From such analysis, we recover a bank of local learned features from a discriminative rule among severity levels of the disease.

An intermediate deep representation is here achieved by implementing receptive field modules that recover an intermediate nodule representation, enhancing the morphological mass features. Hence, the set of convolution activations $\mathbf{A_i}$ are mapped through a Receptive Field Block (RFB), denoted as ($\phi(\mathbf{A^L}) = \mathbf{RFB}(\mathbf{A^L})$). The relative size and texture of the nodules are handled by calculating different kernel sizes, paddings, and dilatations. The high variability of the image is covered by computing multiple responses at different scales in a dense way. This projection represents a multiscale system that outputs several feature maps, from different RFB scales, learning larger neighborhoods and multi-scale textural and shape patterns.

4.2. Multi-head volumetric attention mechanism

Nodule characterization needs to be enriched by recovering non-local textural features, adjusted with respect to the different disease stages. For doing so, multiple self-attention mechanisms are introduced, that complementary recover complex volumetric nodule relationships and enhance discriminative malignancy representations.

Specifically, from the output of the convolutional representation $\mathbf{A^L}$ are coded H self-attention mechanisms $\{\mathbf{SA}^h\}_{h=1}^{\mathbf{H}}$. Each self-attention mechanism is composed of three independent branches that learn an intermediate feature extraction by means of an RFB module, computed as the key ($\mathbf{K}(\phi(\mathbf{A^L})) = \mathbf{W_k}\phi(\mathbf{A^L})$), the query ($\mathbf{Q}(\phi(\mathbf{A^L})) = \mathbf{W_q}\phi(\mathbf{A^L})$) and the value ($\mathbf{V}(\phi(\mathbf{A^L})) = \mathbf{W_v}\phi(\mathbf{A^L})$). A similarity matrix is then recovered from two branches, as $\mathbf{S}_{\overline{ij}}\sigma(\mathbf{K}(\phi(\mathbf{A_i^L}))^T\mathbf{Q}(\phi(\mathbf{A_j^L})))$, where σ represent the softmax fusion. This matrix recovers dense and robust similarities among spatial projections of convolutional decompositions, that code relationships among whole nodule voxels. Then, the value branch is weighted regarding the similarity attention matrix, where attention-weighted features correspond to $\Psi = \mathbf{S} \odot \mathbf{V}(\phi(\mathbf{A^L}))$, enhancing more correlated nodule information.

To enrich the nodule representation, several $\mathbf{SA}$ (multi-head attention) mechanisms are computed in parallel, specialized in extract complementary intermediate and non-local features. The initial adjusting of these mechanisms is challenging, due to the dense correlation among all pixels related within a region of interest. For initial training steps, such representation may propagate strong corrections to the convolutional backbone, which may collapse the nodule learning, from an end-to-end perspective. Then, a set of learnable parameters $\Gamma = \{\gamma^h\}_{h=1}^{\mathbf{H}}$ are defined, which mitigate the learning contribution of attention-mechanisms during initial training epochs and progressively increases the contribution of such representation. The iterative updating $\{\gamma^h\psi^h\}_{h=1}^{\mathbf{H}}$ allows to control the contribution of the attention mechanism during the learning stage.

The final multi-head attention output is described as $\mathbf{SA}^h = \Gamma^h \mathbf{\Psi}^h + \mathbf{A}^{\mathbf{L}}$. The resultant visual descriptor from such a multi-head mechanism allows reconstructing robustly each nodule texture, morphology, and size. The reconstruction is made by different correlation findings presented on the nodule, providing a good estimator of the nodule diagnosis when concatenating each self-attention mechanism. Each attention has a specific delineation and correlation of the nodule, forming together a robust decomposition of the nodule and achieving a dense-feature representation.

4.3. Focal loss

Many of the strategies to support nodule classification are bounded to be binary problems, facilitating to lead with the amount of available data and operating over extreme decisions about the malignancy stage. Nonetheless, in clinical practice, a crucial support for nodule characterization lies in different intermediate stages (moderately benign and malign), which are key to define an early treatment of lung cancer disease. This discrimination among intermediate stages is however challenging even for expert radiologists, being many of these nodules are labeled as indeterminated ($\sim 42.39\%$). These indeterminated nodules present mixed textural patterns, and can only be discriminated with further analyses (second observer, histopathology analysis, or follow-up in a new CT study).

To support an intermediate-stage characterization, computational tools should be addressed to multiple classification strategies with sufficient capability to discriminate among close textural nodule patterns. In fact, learning a robust nodule discriminative representation strongly depends on a well-defined set of labeled data, with a sufficient amount to learn textural variabilities, but also balanced among considered classes to avoid dominant class biases. The main drawback of current multi-classification approaches lies in the cross-entropy minimization rule that easily may be biased into an unbalanced scenario, due to the resultant average error formulation being further back-propagated through the network. To overcome such a problem, the proposed strategy is here penalized with a

focal loss (FL), which allows to focus on sparse samples during training, while considering a down-weighting of the easy positive/negative samples to prevent bias. Mathematically, the focal loss is constructed as $\mathbf{FL}(\mathcal{P}_t) = -(1-\mathcal{P}_t)^{\gamma} log(\mathcal{P}_t)$, where $(1-\mathcal{P}_t)^{\gamma}$ is called the modulating factor, and an hyper-parameter γ is defined as greater or equal to zero. When $\gamma > 0$, allows to reduce the loss contribution on samples that are well classified, and therefore, enables to increase the importance when correcting miss-classified nodule samples.

5. EXPERIMENTAL SETUP

The LIDC-IDRI dataset was used to train and validate the proposed strategy, adopting its malignancy and morphological metadata[63]. This dataset includes 1018 low-dose CT scans, from 1010 patients stratified according to lung cancer severity, observed from characterized nodules. According to Lung-RADS® protocol, each nodule was associated with a specific malignancy diagnosis ranging from $[1-5]$, corresponding to highly benign, moderately benign, indeterminate, moderately malign, and highly malign. Different morphological features were also considered for diagnostic purposes, such as spiculation, size, lobulation, margin, subtlety, and texture, among others. Each feature is also stratified among five levels, except for calcification which also includes the level of absent calcification (level 6).

This study was performed by four experienced thoracic radiologists, who provided the diagnosis, taking into account morphological features and the annotation and delineation of each nodule. For implementation purposes, we considered an agreement made for at least three radiologists, according to the median malignancy value (MMV). This MMV criterion has been widely used in the literature for the validation of the LIDC-IDRI dataset[64] [65]. In such a way, for each nodule, the associated diagnosis labels are ordered in an ascendant way, and then the mid-high label value is assigned to the observation.—1

---

[63] Samuel G Armato III et al. “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans”. In: *Medical physics* 38.2 (2011), pp. 915–931.

[64] Dey, Lu, and Hong, “Diagnostic classification of lung nodules using 3D neural networks”.

[65] Raul Victor M Da Nobrega et al. “Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks”. In: *Neural Computing and Applications* 32 (2020), pp. 11065–11082.

In total, there are 2612 annotated lung nodules, nevertheless, only 887 were annotated among the four radiologists. As seen in Table 2, it can be appreciated that most diagnosis lies on the indeterminate class ($\sim 42.39\%$ of the dataset), where nodules present mixed benign and malign patterns. In contrast, there are few samples of moderately benign nodules ($\sim 5.52\%$), evidencing a highly unbalanced dataset.

Table 2. Amount of nodules classified by each radiologist.

	Highly benign	**Moderately benign**	**Indeterminate**	**Moderately malign**	**Highly malign**
First radiologist	135	196	260	156	140
Second radiologist	130	115	335	162	145
Third radiologist	126	73	349	174	165
Fourth radiologist	142	276	260	146	63
MMV agreement	109	49	376	201	152

5.1. Implementation Details

Each potential nodule observation was cropped from CT studies in a volumetric patch with a size of $32 \times 32 \times 5$. To reduce the influence of unbalanced classes, the dataset was tripled with a data augmentation scheme, by applying horizontal and vertical flipping to each cropped nodule. FL procedure was applied using $\gamma = 5$. For each configuration, two Residual Convolutional blocks (ResConvs) are considered with $3 \times 3 \times 3$ convolutions and a skip-connection. Then, for each branch (key, query, and value), an RFB is calculated with different convolution kernels and dilations $[1, 3, 5, 7]$ to retrieve an intermediate feature representation. For training purposes, an Adam optimization with a learning rate of $1e^{-4}$ was used, together with 50 epochs.

5.2. Validation procedure

A k-fold cross-validation ($k = 10$) was performed for each considered experiment, reporting diverse performance metrics such as precision, sensitivity, AUC, and F1-score. To

consider the AUC, a One-vs-Rest methodology was used to obtain the true positive rate (TPR) ratio with respect to the false positive rate (FPR). Hence, the average mean score was reported for each class. Then, an exhaustive set of experiments were performed to analyze the proposed approach regarding multi-class nodule classification and indeterminate nodules procedure from different perspectives, as follows:

1. **Nodule classification.** In this case, experiments were run with and without the indeterminate nodules. A standard multi-class validation was here considered, where each class has an independent probability output. Also, a one-vs-all scheme was considered during the validation. In such last case, each model gives a binary prediction between a particular class and the rest of the samples. This validation was carried out to establish a comparison with some approaches in the state-of-the-art. Each considered experiment was carried out in a k-fold $(k = 10)$ cross-validation scheme. At each fold, a training $(80\% \sim 511$ nodules) and test $(20\% \sim 52$ nodules) partition was contemplated. For each experiment is guaranteed that there is no patient overlap, following splits from CT studies. Besides, the validation follows a nodule-level strategy, since one patient may have different nodules, but the diagnosis for each one may be different.

2. **Agreement among radiologists.** An agreement study among the four radiologists for each nodule was also considered. This study was performed with a κ value and included the malignancy annotations, but also the stratification of morphological features. The Median Malignancy value (MMV) was adopted as the general score.

3. **Indeterminate nodules support.** The main challenge in nodule characterization is to associate a malignancy degree to certain nodules that share mixed morphological features, which also represent a remarkable uncertainty for cancer diagnosis. Therefore, the indeterminate samples were projected into the trained proposed multi-attention approach to associate them with some malignancy degree. The coherence

of predictions was further validated using a correlation study with other morphological features. Also, the resultant sample distribution was statistically validated with respect to the population of known stratified degrees.

6. EVALUATION AND RESULTS

6.1. Nodule classification

Firstly, an ablation study was carried out by varying different self-attention mechanisms in parallel (heads) ($\{\mathbf{SA}^h\}_{h=1}^4$). Table 3 summarizes achieved results from different classification metrics. In general, the composition with several heads achieves coherent results, providing reliable classifications. To select the best configuration, the general AUC performances were observed at different configurations. This metric allows to generally measure the discrimination at different classification thresholds, allowing to retrieve the most robust configuration. In this case, the best configuration is the one with three heads, achieving an AUC, precision, and sensitivity of 85.35%, 66.81%, and 81.86%, respectively. Despite the four-heads self-attention mechanism having a better chance to specialize filters, training of this mechanism implies a significant computational cost that may introduce instability during learning, thus failing on the representation.

Table 3. Ablation study of multiple self-attention mechanisms in parallel (heads). The second value (in parenthesis) represents a different cut-off during the test.

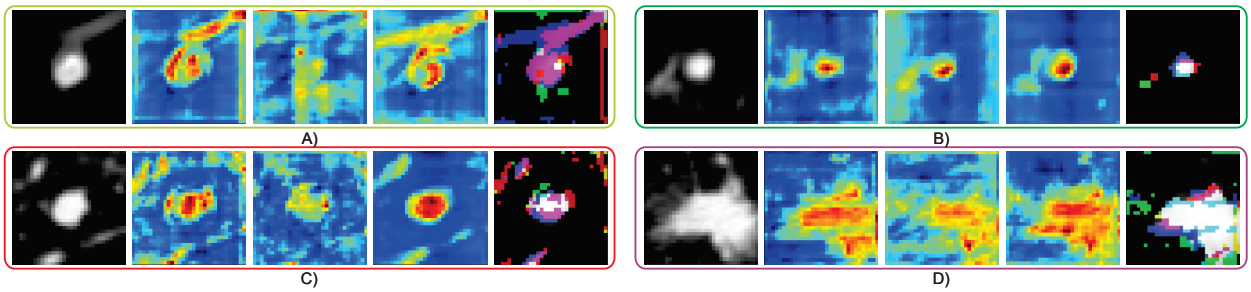
Self-attention mechanism	Acc(%)	Prec(%)	Sens(%)	F1(%)	AUC(%)
1 head	*63.96* **(81.47)**	64.01 **(69.23)**	63.96 (78.83)	62.96 (70.74)	84.09
2 heads	61.89 (80.10)	63.35 (64.76)	61.89 **(81.86)**	60.49 (69.56)	84.00
3 heads	63.78 (80.71)	*64.47 (66.81)*	*63.78 (81.65)*	**63.03 (71.10)**	**85.35**
4 heads	**64.15** *(81.18)*	**65.96** (66.18)	**64.15** (79.47)	62.90 (69.69)	84.63

Hence, following the three-attention heads configuration, the FL and the Cross-Entropy Loss (CEL) minimization rules were compared. The CEL achieves an AUC of 79.40%, while the FL achieves 85.35%. As expected, FL achieves a remarkable performance in the learning process to discriminate malign vs benign nodule stages, even with the small amount of data and weighting classes. This can be due to the fact that FL, in comparison

with CEL, focuses on sparse samples. For doing so, the $\gamma > 0$ allows to down-weight to the easy positive/negative samples to avoid bias in the learning stage.

Secondly, a qualitative analysis was carried out by showing the resultant attention maps from malign and benign texture and morphology features, using the three attention-heads configuration. Figure 11 illustrates nodule samples with their respective attention maps, being the last column a unified version of the nodule attention representation. As can be observed, each attention head learns a specific part of the nodule, enhancing the attention output with local, non-local, and intermediate feature decompositions. Interestingly, for moderately benign nodules (representing $\sim 5.52\%$ of the dataset) the feature maps are properly recovered, focusing the attention on the nodule area. Also for malignant classes, the proposed approach recovered isolated regions with highlighted textures, that may correspond to spiculated/margin features. In general, the recovered output retrieves adequate nodule representations, by learning the variations in texture, noise, size, and morphology.

Figure 11. Examples of feature attention maps obtained for each nodule class: A) highly benign, B) moderately benign, C) moderately malign, D) highly malign. Each image depicts the input nodule, the individual first-head, second-head, and third-head attention maps, and a final concatenation of all attention maps.



6.2. Comparison with the state-of-the-art

A comparison with the state-of-the-art baseline approaches was carried out, following two typical validation methodologies. Firstly, a multi-class validation was proposed, including an experiment where the intermediate nodules are considered as an actual class (as proposed in baselines[66] [67]). Table 4 shows the compiled experiments with the multi-attention architecture: (1) including the indeterminate class with only purposes of state-of-the-art comparison evidencing that for these nodules there is not a consensus among experts, and (2) without the indeterminate class. The proposed approach achieved a remarkable AUC of 78.03% and 85.35%, respectively, in comparison with the state-of-the-art approaches. In fact, the multi-attention architecture presents multiple cut-offs during the test; then, having a different multi-class threshold, we achieved an accuracy of 76.52% and 80.71%, respectively. In such an experiment, it should be noted the dataset imbalance, where the indeterminate nodules represent a massive amount of the available examples, making the classification even more challenging given the mixed (benign/malign) patterns.

Table 4. Comparison of the proposed approach with some state-of-the-art proposals in a multi-class scenario. The second value (in parenthesis) represents different cuts-off during the test.

Method	Acc(%)	Prec(%)	Sens(%)	AUC(%)
Zinovev *et el.*[68]	63.10	-	-	74.60
Kumar *et al.*[69]	*75.01*	-	**83.35**	-
Yung *et al.*[70]	66.36	-	-	-
MA - with Ind class	58.58 **(76.52)**	*55.85 (52.37)*	*58.58 (72.99)*	**78.03**
MA - no Ind class	63.78 **(80.71)**	**64.47 (66.81)**	63.78 *(81.65)*	**85.35**

A complimentary multi-classification was also carried out, following a one-vs-all scheme.

---

[66] Zinovev, Furst, and Raicu, “Building an ensemble of probabilistic classifiers for lung nodule interpretation”.

[67] Yung et al., “Synthetic sampling for multi-class malignancy prediction”.

Table 5 compiles the results and metrics (accuracy, sensitivity, specificity, and AUC) obtained for each of these configurations. As shown there, the proposed multi-attention method outperforms the state-of-the-art for all configurations in terms of sensitivity, by up to 95.35% on benign stages (1, 2 vs all) and 74.24% on malign stages (4, 5 vs all). In fact, benign classes (1, 2) are mostly well classified, in comparison with malignant classes (4, 5). This can be due to the high variability intra-classes among malignant nodules, with different and variable textures, densities, sizes, and morphologies. Nevertheless, the multi-attention method has demonstrated competitive results on the state-of-the-art for all configurations. In general, the proposed approach achieves consistent results along whole considered classes, recovering complementary attention nodule features to achieve a reliable classification. In contrast, Roy *et al.*[71] achieves a better AUC for the highly malignant (5) and moderately benign (2) degrees, but requires updating a generative constraint that may be sensitive to lead with nodule variability. Also, Xie (b) *et al.*[72] demonstrates a better AUC on the moderately malignant (4) degree, using nine views of the nodule and requiring an UNet for each view, thus involving a high amount of parameters.

The proposed approach was also adapted to perform a binary classification, considering two classes: malign (combining moderately malignant and highly malignant) and benign (including moderately benign and highly benign). The proposed volumetric attention approach achieves precision, sensitivity, and AUC values of 90.57%, 91.60%, and 93.95%, respectively, being competitive in comparison with the results obtained by other binary state-of-the-art approaches.

Some studies in the state-of-the-art also report analyses regarding the agreement in the

---

[71] Roy, Mazumdar, and Chowdhury, “ADGAN: Attribute-Driven Generative Adversarial Network for Synthesis and Multiclass Classification of Pulmonary Nodules”.

[72] Xie et al., “Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT”.

Table 5. Comparison of the proposed approach and the state-of-the-art in a one-vs-all scheme.

Class	Methods	Acc(%)	Sens(%)	Spec(%)	AUC(%)
1 vs all	Xie (a) *et al.*[73]	89.62	69.69	94.14	88.94
	Xie (b) *et al.*[74]	*92.95*	73.26	**97.38**	91.57
	Roy *et al.*[75]	**93.7**	*81.4*	*95.6*	*94.1*
	Multi-attention	92.13	[76]	92.13	**95.42**
2 vs all	Xie (a) *et al.*[77]	78.71	83.37	74.33	81.03
	Xie (b) *et al.*[78]	84.12	*89.82*	78.76	*87.76*
	Roy *et al.*[79]	*89.1*	88.7	*82.4*	**91.2**
	Multi-attention	**90.25**	**96.05**	**90.25**	86.05
4 vs all	Xie (a) *et al.*[80]	80.63	46.48	91.64	80.04
	Xie (b) *et al.*[81]	*84.53*	54.19	**94.30**	**89.08**
	Roy *et al.*[82]	**87.3**	*68.1*	*93.2*	*88.7*
	Multi-attention	61.56	**74.24**	61.56	68.59
5 vs all	Xie (a) *et al.*[83]	*90.96*	55.23	94.38	87.30
	Xie (b) *et al.*[84]	**93.72**	67.23	*96.26*	*94.86*
	Roy *et al.*[85]	90.4	*77.1*	**97.3**	**95.4**
	Multi-attention	76.37	**82.76**	76.37	81.53

annotations performed by radiologists. In this work, an agreement study was conducted involving the four radiologists in the dataset, the calculated MMV, and the outputs retrieved by the proposed approach. Figure 12 shows the agreement in each case, evidencing that three of the four radiologists achieve a moderate agreement regarding the MMV. The proposed approach reports also consistent moderate agreement with respect to the MMV, suggesting that this architecture can simulate an expert on the field in order to support the indeterminate nodule diagnosis. Moreover, this analysis is consistent with reported studies in the state-of-the-art that have reported moderate agreement among radiologist experts[86].

[86] Riel et al., "Observer variability for classification of pulmonary nodules on low-dose CT images and its

Figure 12. Malignancy agreements measured with κ values between the radiologist and our model in comparison with the MMV agreement.

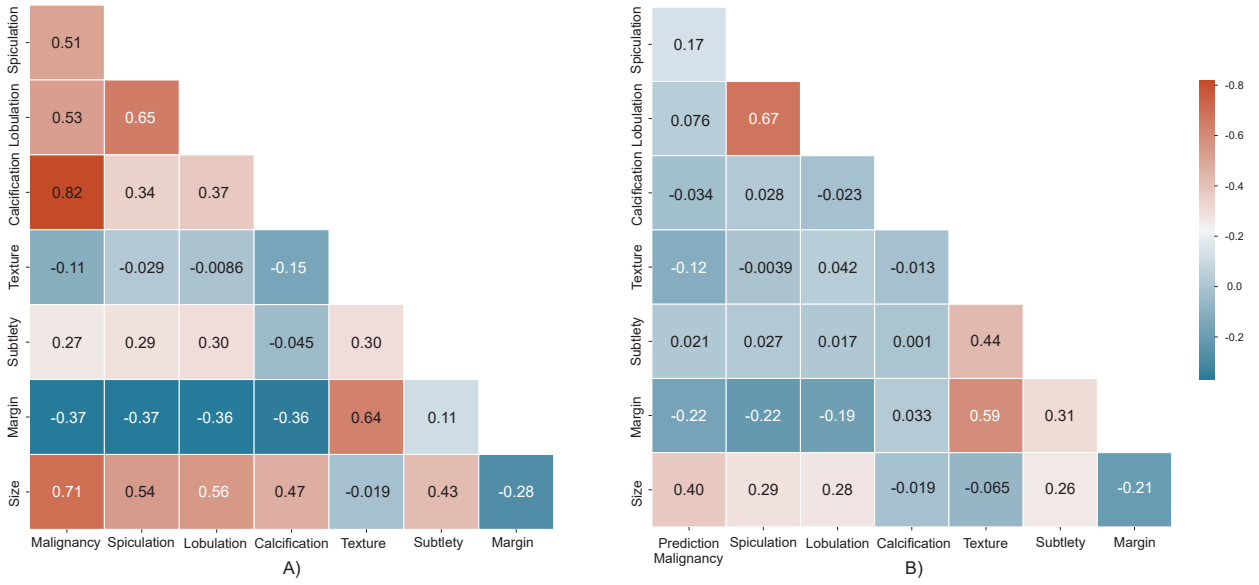
6.3. Exploration of the indeterminate nodules

Indeterminate nodules represent a main diagnosis challenge, hindering a final expert consensus about the disease categorization. Computer-aided diagnosis solutions may be the key in such scenarios, providing additional nodule categorizations or aiming to achieve a consensus among a group of radiologists. To emulate such a task, the resultant indeterminate nodules (376 nodules that correspond to $\sim 42.39\%$ of the complete LIDC-IDRI dataset) were projected to the proposed multi-attention embedding, to associate the most probable class prediction. Thereby, these features may be grouped in pseudo-labels of indeterminated nodules to find coherence regarding the classification. Complementary nodule (morphological) patterns that may be correlated with the disease are further explored, in order to validate the concordance with the label retrieved.

effect on nodule management".

The indeterminate nodules on the LIDC-IDRI dataset, do not have a ground truth or additional information to associate with a nodule class. Hence, we explore the malignancy correlation with other nodule features in a supervised and labeled dataset, to determine some malignancy association with these characteristics. Figure 13-A illustrates the linear correlation from stratified nodules at different levels of the disease. As shown there, a higher correlation is mostly obtained in the malignancy concerning the spiculation (0.51), lobulation (0.53), size (0.71), and calcification (0.82). Also, the margin retrieves a negative correlation (-0.37), having an inverse proportionality; if the nodule presents a sharp morphology, has a benign pattern, but if is poorly defined, presents a malign pattern.

Figure 13. A) Correlation between morphologies: malignancy (highly and moderately malign or benign classes), spiculation, lobulation, calcification, texture, subtlety, margin, and size. B) Indeterminate value inferences described from morphological diagnosis.

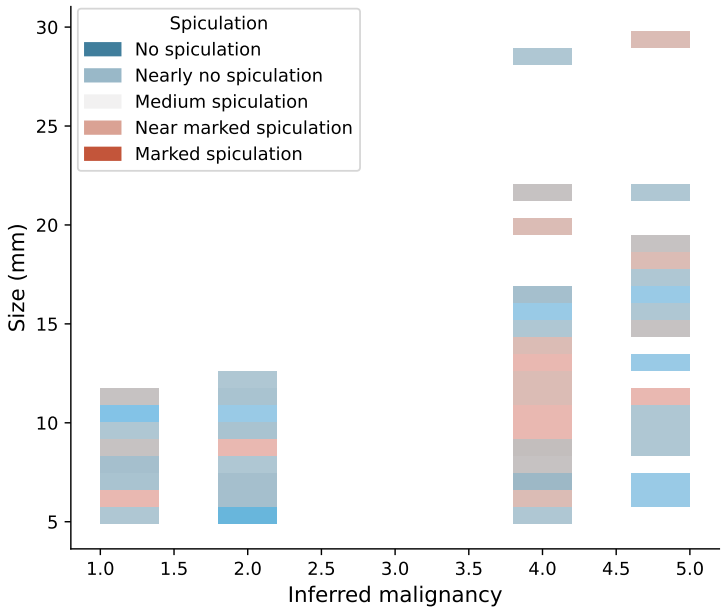


Consequently, Figure 13-B exhibits correlations achieved between the malignancy prediction of the indeterminate nodules (in the proposed embedding) and the corresponding morphological annotations taken from the LIDC-IDRI dataset. It should be noted that the size (0.4) and the margin (-0.22) have in comparison almost the same correlation, re-

trieving good information about the nodules. On the other hand, the spiculation (0.17) and lobulation (0.076) presented a minor value, but have a positive correlation. Although indeterminate nodules have mixed patterns, these morphologies may be crucial to identifying indeterminate nodules and textural differences.

In a detailed analysis, Figure 14 illustrates the respective size, pseudo-stratification, and the corresponding spiculation for the set of classified nodules. As expected, a certain correlation exists between size and malignancy degree, as larger nodules are associated with major malignancy degrees. Additionally, the spiculation is normally detected and generalized to have strong malignancy patterns. Each spiculation value is described in an ascendant manner where 1 corresponds to no spiculation, and 5 represents a marked spiculation. The malign nodule stages present higher size nodules (up to 29.07 mm), and a high level of spiculation, presenting signs of a malignant nodule.

Figure 14. Indeterminate value pseudo-stratification described from size and spiculation.



Finally, a statistical test was also considered here to determine the statistical correspondence of pseudo-labels regarding the actual nodule population with a determined malig-

nancy. A comparison between the stratified and the pseudo-label distributions was then performed to establish a similarity among them. The spiculation distribution of benign and malign (highly and moderately), and the size of benign (highly and moderately) inferences concerning such class retrieves a Mann-Whitney test of $p > 0.05$, showing a non-statistical difference among distributions. On the other hand, regarding the samples labeled in malign stages, the indeterminate size nodules retrieved a $p < 0.005$, which resulted in a non-correspondence with samples of such classes. Such a result may be attributed to the high variability of samples and biased on small nodules stratified with these classes.

7. DISCUSSION

A multi-attention architecture was herein proposed to automatically stratify lung nodules malignancy, observed from low-dose CT images. This multi-attention approach is based on a volumetric representation that retrieves local, intermediate, and non-local textural patterns from a multi-head attention mechanism. The representation enhances the nodule characterization on different fields and scales, capturing dense textural and morphological regions to increase the robust embedding representation. The proposed approach, using three attention-heads, achieved an AUC of 85.35% on the multi-classification task and an average AUC of 82.90% at one-vs-all multi-class schemes.

The proposed approach achieved remarkable sensitivity regarding validation into a one-vs-all validation methodology, reducing the false positive rate in nodule classifications. In such a case, the proposed approach outperformed the SoA in terms of sensitivity with an average of 81.10%. Baseline approaches use redundant information expressed as integrated handcrafted and deep-learning features[87] (sensitivity = 71.13%), U-net auto-encoders[88] (sensitivity = 63.69%), and even also from generative representations, to learn textural nodule patterns[89] (sensitivity = 78.83%). In such cases, the multi-head attention modules achieved an adequate recovery of nodule features to discriminate among different stages of malignancy. In fact, the proposed approach reports a better true positive rate that impacts in sensitivity score, but the discrimination among classes is impacted in

---

87 Xie et al., "Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT".

88 Xie et al., "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT".

89 Roy, Mazumdar, and Chowdhury, "ADGAN: Attribute-Driven Generative Adversarial Network for Synthesis and Multiclass Classification of Pulmonary Nodules".

the AUC. In contrast, the generative scheme[90] reports a better AUC (92.35%) but with a natural increase of false negatives samples.

Particularly in this study, we are interested in the analysis and projection of indeterminate nodules, which typically represent a significant amount of samples in clinical routine. Some approaches in the state-of-the-art had included the indeterminate nodules as an extra class on their respective methodologies. The results may be biased since these nodules provide a non-significant prediction of the clinical context ($\sim 42.39\%$ in the LIDC-IDRI dataset).

In such a case, the approach achieves an AUC and accuracy of 78.03% and 76.52% respectively, while standard approaches achieve an AUC of 74.60% and accuracy values up to 75.01%. These approaches build textural descriptors and discrimination rules from decision trees, AdaBoost[91] and Random Forest strategies[92]. These results may be supported by methodological assumptions of baseline approaches that lost much of the non-local characterization, which impacts a robust quantification of the nodule's dense textural and morphology features. In a more deep analysis, this work introduced new experiments to support the malignancy degree of the indeterminate set of nodules. For doing so, the proposed approach was also included in an agreement expert analysis achieving a moderate agreement regarding MMV and showing a typical radiologist performance. The indeterminated nodules were projected into the proposed multi-attention embedding, obtaining a pseudo-classification among stratified values (highly and moderate malign and benign nodules). Over this study of the LIDC-IDRI dataset, the four radiologists involved in the

---

[90] Roy, Mazumdar, and Chowdhury, "ADGAN: Attribute-Driven Generative Adversarial Network for Synthesis and Multiclass Classification of Pulmonary Nodules".

[91] Zinovev, Furst, and Raicu, "Building an ensemble of probabilistic classifiers for lung nodule interpretation".

[92] Yung et al., "Synthetic sampling for multi-class malignancy prediction".

study achieved moderate agreement. Remarkably, the multi-attention may support the malignancy stratification of the indeterminate nodules almost as well as an average radiologist.

Complementary, the coherence of the retrieved predictions for the proposed approach was validated regarding the correlation between morphology features with the well-defined classification. A proper stratification of indeterminate nodules using size (correlation of 0.4) and spiculation (correlation of 0.17) was then evidenced. Benign pseudo-labels present mostly near non-speculation and small sizes (lower than 15mm) and malignant pseudo-labels have different marked spiculation patterns, and diverse size classes (up to 30mm). From a statistical test analysis, we found that pseudo-labels and stratified benign (highly / moderately) distributions (between malignancy, spiculation, and size) have non-statistical differences ($p > 0.05$), while for malignant size populations, significative differences can be found. The heterogeneity of such samples may be attributed to differences at this stage.

Finally, the proposed approach evidences promising results to support the nodule malignancy classification, according to visual observations. This tool may support analysis during diagnosis, complementing other clinical variables and contributing to reduce subjectivity. Also, from the exposed results over indeterminated nodules, the proposed approach may support malignancy associations, according to previously observed nodules. To transfer such technology in clinical scenarios is now demanding to carry out validations over larger external datasets to measure the generalization capability. Also, there is necessary to perform experiments with nodule observations that include a biopsy or progression stages information, especially for indeterminated masses. Complementary, risk, and mortality information can be approached for the proposed methodology to predict other indexes of the disease.

8. CONCLUSIONS AND FUTURE WORK

In this work a deep representation to classify nodules according to different levels of malignancy is proposed. This representation encodes deep convolutional features in a multi-head architecture to enhance nodule representation and achieve better discrimination among nodule classes. The proposed approach was exhaustively validated in a public dataset, according to the agreement of four radiologists. The achieved results evidence a strong capability to support nodule diagnosis. In a more detailed analysis, the proposed approach was validated according to the classification of indeterminated nodules. In such a case, the predicted classes for such nodules share morphological features with the respective malignancy population. Future works will include a wider study of the indeterminate nodules, including a biopsy and progression stages, in order to provide and support a higher reliable classification. Also, different topological deep mappings could be established in order to discriminate more precisely the malignancy stages between diverse features. Finally, validation with a larger dataset, to explore and robustly classify nodules, should be taken into account, as well as the advantages and limitations of the approach.

BIBLIOGRAPHY

Albert, Ross H and John J Russell. “Evaluation of the solitary pulmonary nodule”. In: *American family physician* 80.8 (2009), pp. 827–831.

Armato III, Samuel G et al. “The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans”. In: *Medical physics* 38.2 (2011), pp. 915–931.

Astaraki, Mehdi et al. “Benign-malignant pulmonary nodule classification in low-dose CT with convolutional features”. In: *Physica Medica* 83 (2021), pp. 146–153.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. “Neural machine translation by jointly learning to align and translate”. In: 2014.

Causey, Jason L et al. “Highly accurate model for prediction of lung nodule malignancy with CT scans”. In: *Scientific reports* 8.1 (2018), pp. 1–12.

Chen, Long et al. “Sca-cnn: Spatial and channel-wise attention in convolutional networks for image captioning”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 5659–5667.

Da Nobrega, Raul Victor M et al. “Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks”. In: *Neural Computing and Applications* 32 (2020), pp. 11065–11082.

Dey, Raunak, Zhongjie Lu, and Yi Hong. “Diagnostic classification of lung nodules using 3D neural networks”. In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE. 2018, pp. 774–778.

Dhara, Ashis Kumar et al. “A combination of shape and texture features for classification of pulmonary nodules in lung CT images”. In: *Journal of digital imaging* 29.4 (2016), pp. 466–475.

Dou, Qi et al. “Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection”. In: *IEEE Transactions on Biomedical Engineering* 64.7 (2016), pp. 1558–1567.

Fu, Xiaohang et al. “Attention-Enhanced Cross-Task Network for Analysing Multiple Attributes of Lung Nodules in CT”. In: 2021.

Han, Fangfang et al. “Texture feature analysis for computer-aided diagnosis on pulmonary nodules”. In: *Journal of digital imaging* 28.1 (2015), pp. 99–115.

Heidarian, Shahin et al. “Cae-transformer: Transformer-based model to predict invasiveness of lung adenocarcinoma subsolid nodules from non-thin section 3d ct scans”. In: 2021.

Jiang, Hanliang et al. “Attentive and ensemble 3D dual path networks for pulmonary nodules classification”. In: *Neurocomputing* 398 (2020), pp. 422–430.

Jiang, Hanliang et al. “Learning efficient, explainable and discriminative representations for pulmonary nodules classification”. In: *Pattern Recognition* 113 (2021), p. 107825.

Kuan, Kingsley et al. “Deep learning for lung cancer detection: tackling the kaggle data science bowl 2017 challenge”. In: *arXiv preprint arXiv:1705.09435* (2017).

Kumar, Devinder, Alexander Wong, and David A Clausi. “Lung nodule classification using deep features in CT images”. In: *2015 12th conference on computer and robot vision*. IEEE. 2015, pp. 133–138.

Liu, Dongxu et al. “Res-trans networks for lung nodule classification”. In: *International Journal of Computer Assisted Radiology and Surgery* (2022), pp. 1–10.

Liu, Hong et al. “Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification”. In: *Journal of Digital Imaging* 33.5 (2020), pp. 1242–1256.

Liu, Songtao, Di Huang, et al. “Receptive field block net for accurate and fast object detection”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 385–400.

Luong, Minh-Thang, Hieu Pham, and Christopher D Manning. “Effective approaches to attention-based neural machine translation”. In: *arXiv preprint arXiv:1508.04025* (2015).

Massion, Pierre P and Ronald C Walker. “Indeterminate pulmonary nodules: risk for having or for developing lung cancer?” In: *Cancer prevention research* 7.12 (2014), pp. 1173–1178.

Mazzone, Peter J and Louis Lam. “Evaluating the patient with a pulmonary nodule: a review”. In: *Jama* 327.3 (2022), pp. 264–273.

McKee, Brady J et al. “Performance of ACR Lung-RADS in a clinical CT lung screening program”. In: *Journal of the American College of Radiology* 13.2 (2016), R25–R29.

McWilliams, Annette et al. “Probability of cancer in pulmonary nodules detected on first screening CT”. In: *New England Journal of Medicine* 369.10 (2013), pp. 910–919.

Organization, World Health. *GLOBOCAN 2020: Cancer Today*. `https://gco.iarc.fr/`. Accessed: 2021-24-08.

Ost, David, Alan M Fein, and Steven H Feinsilver. “The solitary pulmonary nodule”. In: *New England Journal of Medicine* 348.25 (2003), pp. 2535–2542.

Paez, Rafael, Michael N Kammer, and Pierre Massion. “Risk stratification of indeterminate pulmonary nodules”. In: *Current Opinion in Pulmonary Medicine* 27.4 (2021), pp. 240–248.

Pinsky, Paul F et al. “Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment”. In: *Annals of internal medicine* 162.7 (2015), pp. 485–491.

Riel, Sarah J van et al. “Observer variability for classification of pulmonary nodules on low-dose CT images and its effect on nodule management”. In: *Radiology* 277.3 (2015), pp. 863–871.

Rodrigues, Murillo B et al. “Health of things algorithms for malignancy level classification of lung nodules”. In: *IEEE Access* 6 (2018), pp. 18592–18601.

Rodriguez-Canales, Jaime, Edwin Parra-Cuentas, and Ignacio I Wistuba. “Diagnosis and molecular classification of lung cancer”. In: *Lung Cancer* (2016), pp. 25–46.

Roy, Rukhmini, Suparna Mazumdar, and Ananda S Chowdhury. “ADGAN: Attribute-Driven Generative Adversarial Network for Synthesis and Multiclass Classification of Pulmonary Nodules”. In: *IEEE Transactions on Neural Networks and Learning Systems* (2022).

Setio, Arnaud Arindra Adiyoso et al. “Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks”. In: *IEEE transactions on medical imaging* 35.5 (2016), pp. 1160–1169.

Al-Shabi, Mundher, Hwee Kuan Lee, and Maxine Tan. “Gated-dilated networks for lung nodule classification in CT scans”. In: *IEEE Access* 7 (2019), pp. 178827–178838.

Al-Shabi, Mundher et al. “Lung nodule classification using deep local–global networks”. In: *International journal of computer assisted radiology and surgery* 14.10 (2019), pp. 1815–1819.

Siegel, Rebecca L. et al. “Cancer statistics, 2020”. In: *CA: A Cancer Journal for Clinicians* 70.1 (2020), pp. 7–30. DOI: 10.3322/caac.21590. eprint: https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21590. URL: https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21590.

Sun, Lingma et al. “Attention-embedded complementary-stream CNN for false positive reduction in pulmonary nodule detection”. In: vol. 133. Elsevier, 2021, p. 104357.

Sung, Hyuna et al. “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: vol. 71. 3. Wiley Online Library, 2021, pp. 209–249.

Vaswani, Ashish et al. “Attention is all you need”. In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.

Woo, Sanghyun et al. “Cbam: Convolutional block attention module”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.

Xiao, Ning et al. “Ensemble classification for predicting the malignancy level of pulmonary nodules on chest computed tomography images”. In: vol. 20. 1. Spandidos Publications, 2020, pp. 401–408.

Xie, Yutong et al. “Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT”. In: *Information Fusion* 42 (2018), pp. 102–110.

Xie, Yutong et al. “Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT”. In: *IEEE transactions on medical imaging* 38.4 (2018), pp. 991–1004.

Yang, Jiancheng et al. “Relational learning between multiple pulmonary nodules via deep set attention transformers”. In: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2020, pp. 1875–1878.

Yung, Matthew et al. “Synthetic sampling for multi-class malignancy prediction”. In: *arXiv preprint arXiv:1807.02608* (2018).

Zhang, Han et al. “Self-attention generative adversarial networks”. In: *International conference on machine learning*. PMLR. 2019, pp. 7354–7363.

Zhu, Wentao et al. “Deeplung: Deep 3d dual path nets for automated pulmonary nodule detection and classification”. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2018, pp. 673–681.

Zinovev, Dmitriy, Jacob Furst, and Daniela Raicu. “Building an ensemble of probabilistic classifiers for lung nodule interpretation”. In: *2011 10th International Conference on Machine Learning and Applications and Workshops* 2 (2011), pp. 155–161.

*

APPENDIX

appendix A. Academic Products

Journals

- Moreno, A., Rueda, A., & Martinez, F., Lung nodule stratification from CT in a multi-attention deep architecture.
  Status: Submitted in Journal of Computerized Medical Imaging and Graphics.

Conference papers

- Moreno, A., Rueda, A., & Martinez, F. (2022, March). A Multi-Scale Self-Attention Network to Discriminate Pulmonary Nodules. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-4). IEEE.
  Status: Presented.

- Moreno, A., Rueda, A., & Martinez, F. (2023, February). A volumetric multi-head attention strategy for lung nodule classification in CT. In *Medical Imaging* 2022: *Computer-Aided Diagnosis*. SPIE.
  Status: Presented.

- Moreno, A., Olmos, J., Guayacán, L. & Martinez, F. (2023, March). Exploiting multi-head attention maps into a deep Riemannian representation to quantify pulmonary nodules. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (pp. 1-4). IEEE.
  Status: Accepted.

Collaborations

- Moreno, A., Bautista, L., & Martinez, F. Cardiac disease classification from spatio-temporal convolution patterns on cine-MRI sequences.
  Status: Submitted in Journal of Biomédica.

- Moreno, A., Pico, J. A., Bautista, L., Guayacán, L, & Martinez, F. (2022). A fast right ventricle segmentation in cine-MRI from a dense hough representation. *Revista Politécnica*, 18(35), 84-97.
  Status: Submitted.

- Ruano, J., Arcila, J., Romo-Bucheli, D., Vargas, C., Rodríguez, J., Mendoza, Ó., ... & Martínez, F. (2022). Deep learning representations to support COVID-19 diagnosis on CT slices. *Biomédica*, 42(1), 170-183.
  Status: Submitted.

- Moreno, A., Rodríguez, J., & Martínez, F. (2022). Kinematic motion representation in Cine-MRI to support cardiac disease classification. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 10(6), 707-718.
  Status: Submitted.