Artificial intelligence solution to classify pulmonary nodules on CT

D. Blanc, V. Racine, A. Khalil, M. Deloche, J.-A. Broyelle, I. Hammouamri, E. Sinitambirivoutin, M. Fiammante, E. Verdier, T. Besson, et al.

Version of Record: https://www.sciencedirect.com/science/article/pii/S2211568420302497
Manuscript_40d044d85f6cba2add7d36218c224779

Artificial intelligence solution to classify pulmonary nodules on CT

Damien BLANC[a], Victor RACINE[a], Antoine KHALIL[b], Maxime DELOCHE[c], Jean-Armand BROYELLE[c], Ilyass HAMMOUAMRI[c], Emrick SINITAMBIRIVOUTIN[c], Marc FIAMMANTE[d], Emmanuel VERDIER[d], Thibaud BESSON[d], Alexandre SADATE[e], Mathieu LEDERLIN[f], François LAURENT[g], Guillaume CHASSAGNON[h], Gilbert FERRETTI[i], Yann DIASCORN[j], Pierre-Yves BRILLET[k], Lucie CASSAGNES[l], Caroline CARAMELLA[m], Antoine LOUBET[n], Neelem ABASSEBAY[o], Philippe CUINGNET[o], Mickael OHANA[p], Julien BEHR[q], Angeline GINZAC[r,s,t], Hugo VEYSSIERE[r,s,t], Xavier DURANDO[r,s,t,u], Imad BOUSAÏD[v,] Nathalie LASSAU[w], Julien BREHANT[x]

Affiliations

a. QuantaCell, IRMB, Hôpital Saint Eloi, 80 Avenue Augustin Fliche, 34090 Montpellier, France
b. Department of Imaging, Neuroradiology unit, Assistance Publique-Hôpitaux de Paris (APHP), Hôpital Bichat Claude Bernard, 75018 Paris, France
c. IBM Cognitive Systems Lab, 34000 Montpellier, France
d. IBM Cognitive Systems France, 92270 Bois-Colombes, France
e. Department of Radiology and medical imaging, CHU Nîmes, Univ. Montpellier, EA2415, 30029 Nîmes, France
f. Department of Radiology, Hôpital universitaire Pontchaillou, 35000 Rennes, France
g. Department of thoracic and cardiovascular Imaging, Respiratory Diseases Service, Respiratory Functional Exploration Service, Hôpital universitaire de Bordeaux, CIC 1401, F-33600, Pessac, France
h. Department of radiology, Hôpital Cochin, APHP. Université de Paris, 27 rue du Faubourg Saint Jacques, 75014 Paris, France
i. Department of Radiology and medical imaging, CHU Grenoble Alpes, 38700 Grenoble, France
j. Department of radiology, Hôpital Universitaire Pasteur, Nice, France
k. INSERM UMR 1272, Université Sorbonne Paris Nord, Bobigny, France; AP-HP, Department of radiology, Hôpital Avicenne, 93430 Bobigny, France
l. Department of radiology B, CHU Gabriel Montpied, 63003 Clermont-Ferrand, France
m. Department of radiology, Institut Gustave ROUSSY, 94800 Villejuif, France
n. Department of neuroradiology, Hôpital Gui-de-Chauliac, CHRU de Montpellier, 34000 Montpellier, France
o. Department of radiology, CH Douai, 59507 Douai, France
p. Department of radiology, Nouvel Hôpital Civil, 67000 Strasbourg, France
q. Department of radiology, CHRU de Jean-Minjoz Besançon, 25030 Besançon, France
r. Clinical Research Unit, Clinical Research and Innovation Delegation, Centre de Lutte contre le Cancer, Centre Jean PERRIN, 58, rue Montalembert, 63011, Clermont-Ferrand Cedex 1, France

s. Université Clermont Auvergne, INSERM, U1240 Imagerie Moléculaire et Stratégies Théranostiques, Centre Jean PERRIN, 63011, Clermont-Ferrand, France
t. Clinical Investigation Centre, UMR501, 63011, Clermont-Ferrand, France
u. Department of medical oncology, Centre Jean PERRIN, 58, rue Montalembert, 63011 Clermont-Ferrand, France
v. Digital Transformation and Information Systems Division, Gustave Roussy, 94800 Villejuif, France.
w. Multimodal Biomedical Imaging Laboratory Paris-Saclay, BIOMAPS, UMR 1281, Université Paris-Saclay, Inserm, CNRS, CEA, 94800 Villejuif, France; Département d'imagerie, Institut Gustave Roussy, 94800 Villejuif, France
x. Department of radiology, Centre Jean PERRIN, 58, rue Montalembert, 63011 Clermont-Ferrand, France

Declaration of interest: None

Corresponding author:
Julien BREHANT
Julien.brehant@clermont.unicancer.fr

Abstract

Purpose: The purpose of this study was to create an algorithm to detect and classify pulmonary nodules in two categories based on their volume greater than 100 mm^3 or not, using machine learning and deep learning techniques.

Materials and methods: The dataset used to train the model was provided by the organization team of the SFR (French Radiological Society) Data Challenge 2019. An asynchronous and parallel 3-stages pipeline was developed to process all the data (a data “pre-processing” stage; a "nodule detection" stage; a “classifier” stage). Lung segmentation was achieved using 3D U-NET algorithm; nodule detection was done using 3D Retina-UNET and classifier stage with a support vector machine algorithm on selected features. Performances were assessed using area under receiver operating characteristics curve (AUROC).

Results: The pipeline showed good performance for pathological nodule detection and patient diagnosis. With the preparation dataset, an AUROC of 0.9058 (95% confidence interval [CI]: 0.8746-0.9362) was obtained, 87% yielding accuracy (95% CI: 84.83%-91.03%) for the "nodule detection" stage, corresponding to 86% specificity (95% CI: 82%-92%) and 89% sensitivity (95% CI: 84.83%-91.03%).

Conclusion: A fully functional pipeline using 3D U-NET, 3D Retina-UNET and classifier stage with a support vector machine algorithm was developed, resulting in high capabilities for pulmonary nodule classification.

Keywords: Lung cancer; Pulmonary nodule; Support vector machine; Deep learning; Machine learning.

List of abbreviations

2D: two-dimensional;

3D: Three-dimensional;

AI: Artificial intelligence;

AUC: Area under the curve;

AUROC: area under receiver operating characteristics curve;

CAD: Computer-aided diagnostic;

CNN: convolutional neural network;

CPU: Central processing units;

CT: Computed tomography;

FN: False negative;

FP: False positive;

GPU: Graphics processing units;

HU: Hounsfield unit;

LIDC: Lung Image Database Consortium;

R-CNN: Region-based convolutional neural network;

RFE: Recursive feature elimination;

ROC: Receiver operating characteristics;

SVM: Support vector machine

Introduction

Lung cancer has the highest mortality rate among all malignant tumors [1]. It is an aggressive disease carrying a dismal prognosis with a 5-year survival rate at 18%. In France, lung cancer remains a major public health problem because of its frequency, but especially the severity of the disease [2]. The key to reduce lung cancer mortality is the accurate diagnosis of pulmonary nodules at early-stage. Early detection and examination of lung nodules, which might be malignant, is necessary [3].

The National Lung Screening Trial study found in more than 53,000 (former-) heavy smokers that annual screening with low-dose computed tomography (CT) reduced lung cancer mortality and overall mortality by 20% and 6.7% respectively [4]. Screening studies using semi-automated volume measurements have shown higher accuracy and reproducibility compared to diameter measurements, and it has been shown that small nodules (those with a volume < 100 mm^3 or diameter < 5 mm) are not predictive for lung cancer [5]. Radiologists spend countless hours carefully detecting small spherical-shaped nodules on CT images [6].

Artificial intelligence (AI) algorithms have demonstrated remarkable progress in image-recognition tasks. Methods ranging from convolutional neural networks (CNN) to variational auto encoders have found myriad applications in the field of medical image

analysis, propelling it forward at a rapid pace [7–11]. Computer-aided diagnostic (CAD) techniques are considered to have potential beyond human experts for accurate diagnosis of early pulmonary nodules. The sensitivity of all the selected works' ability to detect pulmonary nodule ranges from 68.9% to 100%, and the false positive rates range from 0.138 to 38.8 per CT examination [12]. Some selected works achieved a sensitivity of greater than 90%, however, high sensitivity is usually associated with a high false positive rate. The CAD system has inherent limitations concerning the detection of advanced lung tumors, and the human eye is still necessary in these patients. Most recently, conventional CAD solutions that require visual confirmation to reduce false-positive calls [13] are being challenged by deep learning algorithms that have an inherent advantage of automatic feature exploitation [14]. The detection and classification of pulmonary nodules based on deep learning technology can continuously improve the accuracy of diagnosis through self-learning, and is an important means to achieve CAD [15].

The purpose of this study was to create an algorithm to detect and classify pulmonary nodules in two categories based on their volumes greater than 100 mm^3 or not, using machine learning and deep learning techniques.

Material and methods

This study has been entirely designed following the General Data Protection Regulation, as ensured by the French regulation office. Members of the AI groups who provided CT examinations have signed an ethical charter. All images were anonymized and pseudonymized.

Dataset

The dataset used to train the model was provided by the Data Challenge 2019 group. It was composed by 1056 * 3D-DICOM CT images, with various resolutions and size (from 512 × 512 × 130 to 512 × 512 × 1300). This dataset access was provided in three batches: 343 DICOM patients with their related nodules annotation for the first month of development, 344 DICOM patients with annotation 3 days before the final and 344 DICOM patients without annotation for the 2 hours final round. The whole dataset size was about 243 Gigabytes in total.

The methods for annotation were not consistent across the dataset. The location and size of the nodules were labelled as:

- a mark of a single voxel in the middle of the nodule
- a larger spot in the middle of nodules on one single two-dimensional (2D) CT slice
- a circle surrounding nodules on one single 2D CT slice
- a 3D volume to segment the selected nodules
- A few DICOM also annotated nodules smaller than 100 mm³ while it was below the challenge detection criteria.

Processing pipeline

Our solution was based on a three-stage multithreaded pipeline implementing inter-steps communication by asynchronous message passing (Figure 1). The first stage, named "pre-processing", uniformed the DICOM data and generated lung segmentation, the second stage, named "deep learning", detected potential nodules, the third stage, named "classification", to classify nodules larger than 100 mm³ and classify if the patient had abnormal findings or not based on selected nodules features.

An orchestration mechanism managed the whole pipeline data workflow and supervised the underlying daemon processes. The main function of this scheduler was to parallelize across our 3-node cluster: efficiently dispatch images across the available resources (central processing units [CPUs] and graphics processing units [GPUs]) for parallel processing. The scheduler was based on bash scripts (Linux *flock* locking system) to ensure each file was only treated once by a job. It maintained a list of all the jobs that were completed, running or pending: a pre-processing job and an inference job for each image in the test dataset. It then checked availability of resources on the worker nodes and scheduled jobs on these resources when idle (through SSH connections to the worker nodes).

The pipeline was developed using a single IBM Power AC922 server offering 22 IBM POWER9 cores, 512 GB of system memory,4 Volta V100 NVIDIA GPU with 32 GB of graphical memory. This server was purposely designed for machine learning and deep learning tasks, providing good performance and robustness. We used additional machine for the final round of the data challenge.

Due to time constraint a parallelized processing by deploying the solution on a cluster of three IBM AC922 for a total for 60 POWER9 cores and 14 Volta 100 GPUs. The interconnection between the jobs was achieved using a fast networking interconnection

technology (InfiniBand EDR switch and adapters). The spectrum scale file system had three advantages: first, it removed the need to perform file transfers that was time consuming for larger 3D CT datasets. Second, it ensured consistency of data. Third it provided sustained high performances required during the challenge and better than network file system.

Preprocessing

The pre-processing phases consist in images normalization and lungs segmentation. CT intensities (in Hounsfield unit [HU]) were clipped between -1200 and 600 to keep air, lungs up to bones densities. Lung segmentation was obtained with a 3D U-NET and CT intensities were normalized to HU.

Pre-processing is a powerful way to improve the accuracy of the nodule detection model, it helps remove unnecessary data information, by allowing the next pipeline stages to focus on relevant features for their task. Thus, we implemented a complex preprocessing stage for nodule region proposal network, mostly inspired from state of the art techniques. Lung mask was computed using a 3D semantic segmentation CNN model. A last scan intensity normalization step was performed to end the pre-processing phase.

Lung segmentation using a 3D CNN model

The aim of the lung segmentation was to reduce the volume to explore for the next 'nodule detection' pipeline stage and to improve the overall response time and avoid wrong detections of nodules outside lungs. It was implemented using a 3D CNN U-Net model and standards Python Computer Vision Techniques, provided by packages such as Pydicom, SimpleITK, nibabel and OpenCV. We have also used the IBM AC922 technology to train this network, which allowed increasing image resolution of the training dataset by 256^3 instead of 128^3, which was initial the limit related to 32 gigabits graphical processing unit memory of NVIDIA V100. Once the model has inferred the lungs segmentation map from the resampled scan, a morphological closing was applied to the model's output in order to prevent eventual holes in the lungs (Figures 2, 3)

CT data transformation

The goal of this phase was to clean and normalize the initial 3D image and also remove unnecessary information for the next pipeline stage (Figure 4). Every processed image was then written on spectrum scale shared file system for robustness in case of failure.

A completion signal was sent to the scheduler, which marked the inference phase for the image as ready.

We resized the 878 CT data sets from Lung Image Database Consortium (LIDC) data to a pixel size of 1.4 × 0.7 × 0.7 mm^3. The LIDC dataset were split in 80/20, giving 700 patients for training, and 178 for validation. For each batch generated, patients were randomly picked and kept according to the presence or absence of one or more nodules in the corresponding scan. Then, for each chosen patient, a random crop of size 128 × 128 × 64 was performed. Data augmentation operations were then applied while respecting a constraint of distance between nodules and the edge of the image. The complete training of the algorithm lasted during 400 epochs. Each of the epochs corresponded to 150 iterations of the batch generation process. Since the network outputs were of different natures, the Retina UNet was trained using a combination of loss used for both segmentation and object detection problem.

Nodule detection

A deep learning pipeline was setup to identify nodules candidates for each DICOM dataset. Each candidates' regions have been classified to determine whether it contains a nodule with a volume greater than 100 mm^3. We used a Retina-UNet model that showed good performance for object detection [16]. This network takes advantage of the U-Net architecture and has been transferred successfully into 3D [17,18] and cancer detection (Figure 5). The network architecture is described in Figure 2. Most important meta-parameters are the patch size taken for tiling: 128 × 128 × 64. The U-shape of the algorithm has 5 steps (ResBlock) for down sampling then up-sampling. Models are trained using Adam optimizer with a 10^-4 learning rate. For each batch generated, patients were randomly picked and kept according to the presence or absence of one or more nodules in the corresponding scan. Then, for each chosen patient, a random crop of size 128 x 128 x 64 is performed with a constraint of balancing crop without nodules (50%) and crop with nodules (50%).

Since the network outputs are of different natures, the Retina Unet was trained using a combination of loss used for both segmentation and object detection problem:

$$\mathscr{L}_{Glob} = \mathscr{L}_{CE_{Box}} + \mathscr{L}_{Huber} + \frac{\mathscr{L}_{CE_{Pix}} + \mathscr{L}_{Dice}}{2}$$

LatexCommand : $\mathscr{L}_{Glob} = \mathscr{L}_{CE_{Box}} +\mathscr{L}_{Huber} + \frac{\mathscr{L}_{CE_{Pix}}+\mathscr{L}_{Dice}}{2}$

Where:

- L_{CE_{Box}}(resp. L_{CE_{Pix}}): is the cross-entropy function for anchor boxes classification (resp. pixel classification). Minimizing this function is essential to improve the quality of the anchor box classifier (and pixel classification).
- \mathscr{L}_{Huber} is the Huber loss function and reflects the distance between the true box coordinates and the predicted ones [19].
- \mathscr{L}_{Dice}} is the Dice loss function: this loss allows penalizing false positive pixel classification and is specially used for medical application [20].

Feature extraction and nodule classification

Nodule classification was only applied on Data Challenge data. Thirty-nine features were extracted from detected 3D regions (Figure 5):

- Probability given by the deep learning detection (one feature)
- Bounding box center, size and volume (7 features)
- Minimum, maximum and median of pixel intensities (3 features)
- For 7 different 3D segmentations: ratio of pixel above the threshold, volume of pixels above the threshold, ratio between average intensity above and below the threshold, number of objects above the threshold. 6 segmentations correspond to 6 intensity thresholds (named T0 to T5) uniformly spread between minimum and maximum intensity values and the 6^{th} segmentation is the segmentation obtained in the deep learning step ($4 \times 7 = 28$ features)

The SVM allowed classify the status of patients in the Data Challenge. To classify positive nodules according to the annotation, a support vector machine (SVM) classifier was used and the posterior probabilities for each nodule were extracted. A linear kernel was used in order to avoid overfitting. To measure the performance of the classifier, a cross validation (5 folds) procedure was setup (Figure 6). Feature selection algorithm was used to eliminate unnecessary features. Recursive feature elimination (RFE) was used [21]. Area under ROC curve (AUROC) was used as performance metric over the cross validation, considering that AUROC indicates general performance of the classifier [23]. The confident intervals were calculated using a 10-fold cross validation.

Results

During the Data Challenge, we had two hours to process 344 new patients, including extracting the ZIP archive containing the dataset. The workload itself was mostly hold by CPUs units. As each AC922 server had 36 physical cores, hence allowing to pre-process more than 100 images in parallel on three servers. Pre-processing required 3.5 min per DICOM, mostly used for lung segmentation. Data heterogeneity and data format were handled using dedicated DICOM and annotation loader. The training of the deep learning step lasted 4 days. The training process was stopped earlier in order to respect time constraint. For object detection problem, both classification and localization of bounding boxes outputs needed to be evaluated. The best model used for the challenge was selected by maximizing the average precision up to 0.87 [22].

The AUROC for patient classification (based on linear SVM) using all 39 features was 0.8811. To optimize it, a feature selection algorithm was used: recursive feature elimination (RFE) [21]. The AUC was maximized at 0.9022 after removing 30 features. Ten features were kept with 1 feature giving the probability of presence from the deep learning algorithm, 2 features describing the width of the nodule candidate, 2 features describing the UNet Mask Segmentation and the 5-remaining described the segmentation at different thresholds. On 10 remaining features, several optimizations were applied, leading to an AUROC of 0.9058. First a fixed threshold T was applied to remove nodule candidates with a probability score lower than the threshold. Then we defined a weight for the two classes in the SVM parameter to 1 for the benign nodules and C for the malignant one. Both parameters T and C were optimized using force brut strategies learning to optimal parameters T = 0.1 and C = 2.1. Finally, we tried to improve the AUROC by increasing the order of the polynomial kernel up to 2 to 5 while it was originally set to 1 (1 corresponding to a linear kernel). We found that increasing the order did not improve AUROC, and then we kept 1 as kernel order.

Finally, we obtain an AUROC of 0.9058 (95% confidence interval [CI]: 0.8746-0.9362). Considering false negative and false positive findings as equally problematic, the specificity was 86% (95% CI: 82%-92%), sensitivity 89% (95% CI: 88.70-92.72%) and accuracy 87% (95% CI: 84.83%-91.03%). On the 687 patients, number of true positives was 301, number of false-negatives was 43, number of true negatives was 297 and number of false-positives was 46. We found 4 non-solid or part-solid pulmonary nodules and 3 were accurately detected while one was missed.

Discussion

In four weeks, we have built a fully functional solution leveraging open source frameworks, algorithms to win the Data Challenge 2019. This study shows good performances to detect pulmonary nodules greater than 100 mm^3 in a very limited time. Based on observations in high-risk patients from lung cancer screening trials a cut diameter below 6 mm (100 mm^3) is proposed by most recent guidelines as an indicator of acceptably low cancer risk (< 1%). AI can help us to tackle this challenge by creating an "augmented radiologist". Machine-learning classifiers are used to predict the malignance and after deep learning, especially the convolutional neural network, who might be suitable for pulmonary nodule classification when characterized by small medical datasets and small targets [15,24-27].

This study gives good performances in terms of all metrics measured including AUROC (0.9058), sensitivity (89%) and specificity (86%). The developed system presents excellent performance using various databases coming from different hospitals, and with different data acquisition protocols. An improved approach might have been to provide more annotated data, or using GPU with larger memory. With larger memory size, tile size can be larger, and more spatial information can be captured by deep learning algorithms. We did not manage to train correctly the network with tile size of 256 × 256 × 64. With that tile size, we could not load more than 4 tiles on the GPU in the network. With such small batch size, the training procedure would have converged too slowly, but this will probably be possible with GPUs next generation. In this study we have used tile size of 128 × 128 × 64 tiles with a batch size of 26. A larger batch size would allow reducing the iteration noise and would converge faster. Detection can be improved by segmenting more precisely the lung because some errors rise from true nodules localized on the pleura leading the preprocessing algorithm to consider the nodule outside of the lung. Another way to improve the system is the classification of the nodule candidates. In this study we used machine learning technique because the number of patients was low. With a larger dataset we would be able to use 3D deep-learning system classifiers that show good performance in 3D medical imaging [28,29]. Finally, some improvements can be obtained by focusing on some situation leading to errors like pulmonary condensations or non-solid and part-solid pulmonary nodules. Those situations should be considered as additional classes and separated for the rest of nodules such as their confusion is decreased [30]. To meet the schedule of the JFR data Challenge, we could manually segment the images in the JFR datasets. Therefore, we decided to use the LUNA16- LIDC DataSet for deep learning model training to detect nodules [31]. LIDC dataset is similar to

JFR one but with high quality segmentation nodules. However, we used the JFR dataset to learn the final status of the patients.

Researchers have developed an AI scheme to predict the invasive degree of ground-glass nodules by using a residual learning architecture [31]. They showed that deep learning scheme yielded equivalent or even higher performance when compared with the radiologists, and residual learning–based CNN model could improve the performance in classifying between invasive adenocarcinoma and non- invasive adenocarcinoma nodules [31]. Radiomics is also useful to detect lung nodules, and allows examining large volumes of complex data to identify precise characteristics or outcomes. In the setting of lung nodules and lung cancer, radiomics is aimed at deriving automated quantitative imaging features that can predict nodule and tumor behavior non-invasively [32].

Our study has some limitations. First, we were only interested in nodules greater than 100 mm3. The quality of the deep learning detection is highly dependent on the diversity learned in the training dataset, and then nodules will be correctly detected only if closed examples are found in the training dataset.

In conclusion, our results are very encouraging, in particular for a precise and fast detection of pulmonary nodules using CT examination. This will need further cooperation between engineers specializing in deep learning, data scientists and radiologists to build solutions suitable to clinical practice.

Acknowledgments

The authors acknowledge all radiologists who supplied chest CT examinations for the data challenge and the Clinical Research Department of Jean PERRIN Cente.

Human rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patients.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure of interest

The authors declare that they have no competing interest.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

References

[1] Chassagnon G, Revel MP. Lung cancer screening: Current status and perspective. Diagn Interv Imaging 2016;97:949–53.

[2] Gounant V, Khalil A, Créquit P, Lavole A, Ruppert AM, Antoine M, et al. 2014 update on non-small cell lung cancer (excluding diagnosis). Diagn Interv Imaging 2014;95:721–5.

[3] Winer-Muram HT. The solitary pulmonary nodule. Radiology 2006;239:34–49.

[4] Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the Two Incidence Screenings in the National Lung Screening Trial. N Engl J Med 2013;369:920–31.

[5] Horeweg N, van Rosmalen J, Heuvelmans MA, van der Aalst CM, Vliegenthart R, Scholten ET, et al. Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. Lancet Oncol 2014;15:1332–41.

[6] Jung H, Kim B, Lee I, Lee J, Kang J. Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method. BMC Med Imaging 2018;18:48.

[7] Lassau N, Bousaid I, Chouzenoux E, Lamarque JP, Charmettant B, Azoulay M, et al. Three artificial intelligence data challenges based on CT and MRI. Diagn Interv Imaging 2020; doi:10.1016/j.diii.2020.03.006.

[8] Beregi J-P, Zins M, Masson J-P, Cart P, Bartoli J-M, Silberman B, et al. Radiology and artificial intelligence: An opportunity for our specialty. Diagn Interv Imaging 2018;99:677–8.

[9] Waymel Q, Badr S, Demondion X, Cotten A, Jacques T. Impact of the rise of artificial intelligence in radiology: what do radiologists think? Diagn Interv Imaging 2019;100:327–36.

[10] Lassau N, Estienne T, de Vomecourt P, Azoulay M, Cagnol J, Garcia G, et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. Diagn Interv Imaging 2019;100:199–209.

[11] Artificial intelligence and medical imaging 2018: French Radiology Community white paper. Diagn Interv Imaging 2018;99:727–42.

[12] Zhang G, Jiang S, Yang Z, Gong L, Ma X, Zhou Z, et al. Automatic nodule detection for lung cancer in CT images: a review. Comput Biol Med 2018;103:287–300.

[13] National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med 2011;365:395–409.

[14] International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 2006;355:1763–71.

[15] Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W. Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg 2018;13:585–95.

[16] Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell 2020;42:318–27.

[17] Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network. IEEE Trans Neural Netw Learn Syst 2019;30:3484–95..

[18] Jaeger PF, Kohl SAA, Bickelhaupt S, Isensee F, Kuder TA, Schlemmer HP, et al. Retina U-net: embarrassingly simple exploitation of segmentation supervision for medical object detection. PMLR 2019;116:171–83.

[19] Huber PJ. Robust estimation of a location parameter. Ann Math Statist 1964;35:73–101.

[20] Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. IEEE Transactions on Medical Imaging 1994;13:716–24..

[21] Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics 2018;19:432.

[22] Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int J Comput Vis 2010;88:303–38.

[23] Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg 2017;12:1799–808.

[24] Baldwin DR, Gustafson J, Pickup L, Arteta C, Novotny P, Declerck J, et al. External validation of a convolutional neural network artificial intelligence tool to predict malignancy in pulmonary nodules. Thorax 2020;75:306–12.

[25] Li D, Mikela VB, Carlsen FJ, Albrecht-Beste E, Lauridsen C, Bachmann Nielsen M, et al. The performance of deep learning algorithms on automatic pulmonary nodule detection and classification tested on different datasets that are not derived from LIDC-IDRI: a systematic review. Diagnostics 2019;9.

[26] El-Bana S, Al-Kabbany A, Sharkas M. A two-stage framework for automated malignant pulmonary nodule detection in CT scans. Diagnostics 2020; doi:10.3390/diagnostics10030131.

[27] González Maldonado S, Delorme S, Hüsing A, Motsch E, Kauczor H-U, Heussel CP, et al. Evaluation of prediction models for identifying malignancy in pulmonary nodules detected via low-dose computed tomography. JAMA Netw Open 2020;3:e1921221.

[28] Oh K, Chung YC, Kim KW, Kim WS, Oh IS. Classification and visualization of Alzheimer disease using volumetric convolutional neural network and transfer learning. Sci Rep 2019;9:18150.

[29] Feng C, Elazab A, Yang P, Wang T, Zhou F, Hu H, et al. Deep learning framework for Alzheimer's disease diagnosis via 3D-CNN and FSBi-LSTM. IEEE Access 2019;7:63605–18.

[30] Tu X, Xie M, Gao J, Ma Z, Chen D, Wang Q, et al. Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in CT images with convolutional neural network. Sci Rep 2017;7:8533.

[31] Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C van den, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 2017;42:1–13.

[31] Gong J, Liu J, Hao W, Nie S, Zheng B, Wang S, et al. A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. Eur Radiol 2020;30:1847–55.

[32] Wilson R, Devaraj A. Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res 2017;6:86–91.

FIGURE CAPTIONS

Figure 1. Figure shows proposed solution as a 3-stage pipeline.

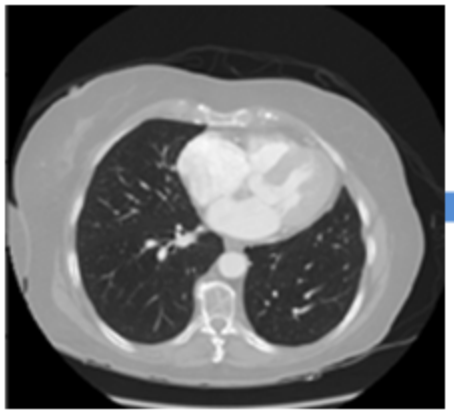
Figure 2. Figure shows effect of morphological closing on lung mask

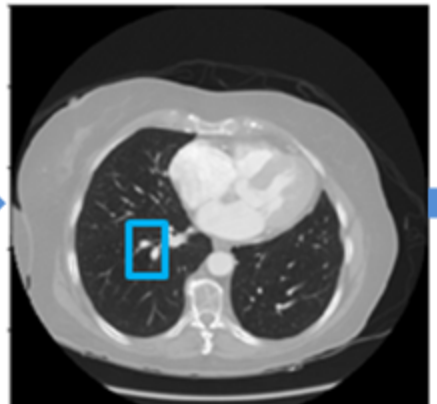
Figure 3. Figure shows segmentation process from CT data to obtain three-dimensional 3D view of lungs in an individual patient.

Figure 4. Figure shows iterative steps of pre-processing. The initial scan intensities (A) are loaded and normalized in Hounsfield Unit (HU). They are resampled to match a given slice thickness that correspond roughly to the average slice thickness of the dataset and are clipped within a given range to limit the impact of extreme values. Then the lung mask (B) is dilated (C) and every pixel outside this dilated version is set to some padding value matching the HU intensity of a normal tissue (C) to remove everything outside the patient's lungs, yet keeping the context that characterizes the nodules at the edges of the lungs. Finally, remaining bones (E) are removed using voxel intensity thresholds (based on the known HU range for bones) in the dilated lung mask version and padded to the same HU value matching normal tissues (F).

Figure 5. Figure shows retina-UNet architecture. Retina-UNet architecture consists of different unified blocks of networks and subnetworks. The block (A) is a Feature Pyramid Network (FPN) based on ResNet architecture and it composes of a bottom-up (A) and a top-down pathway. The bottom-up pathway allows feature extraction while decreasing successively the features maps resolution. Then the top-down pathway applies an upsampling operation to each layer and merges it with the previous layer. The last layer of the top-down pathway is used to output a segmentation mask (C) while less resolved layers generate the features maps inputs for the second block used for object detection (C). Boxes are detected by applying block (D) at different resolution. During this stage, anchor boxes of different size are generated on the input feature map. Regions of interest correspond to a positive bounding box with shift, computed through an anchor boxes classifier and an anchor boxes coordinates regressor. Block (E) details operations represented by different arrows.

Figure 6. Figure shows feature extraction and nodule classification. The patient based cross validation is showed in A. AUC calculation from a patient list in B. The feature selection process with RFE showing the iterative in blue in C. The X labels correspond (from left to right) to the ordered list of features iteratively removed. Thus, the first feature removed is InitScoreFeature, and then BBCenterX. The red point in C corresponds to the optimal RFE iteration. This iteration corresponds to the removal of the features located on the red point's left: features from InitScoreFeature up to BBWidthY are removed. In D, the different feature optimization are displayed, starting with the full set of features (InitScoreFeature), then set of feature after RFE (ScoreFeatureSelection), then optimization of the weight and threshold on low probability (ScoreBestThreshold&Weight) and finally the optimization on the polynomial order (ScoreBestPolyOrder).





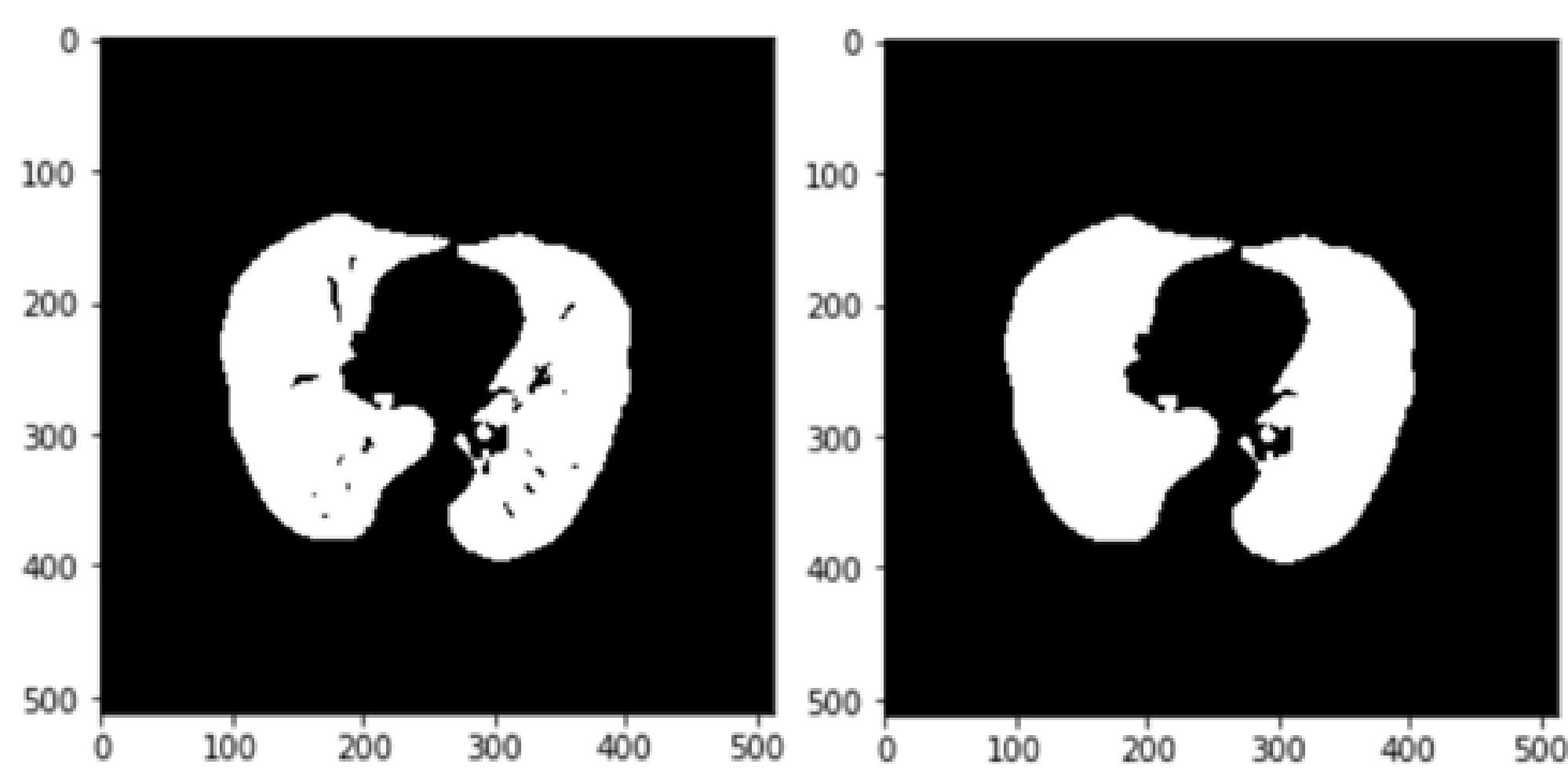
Patient_1 Abnormal

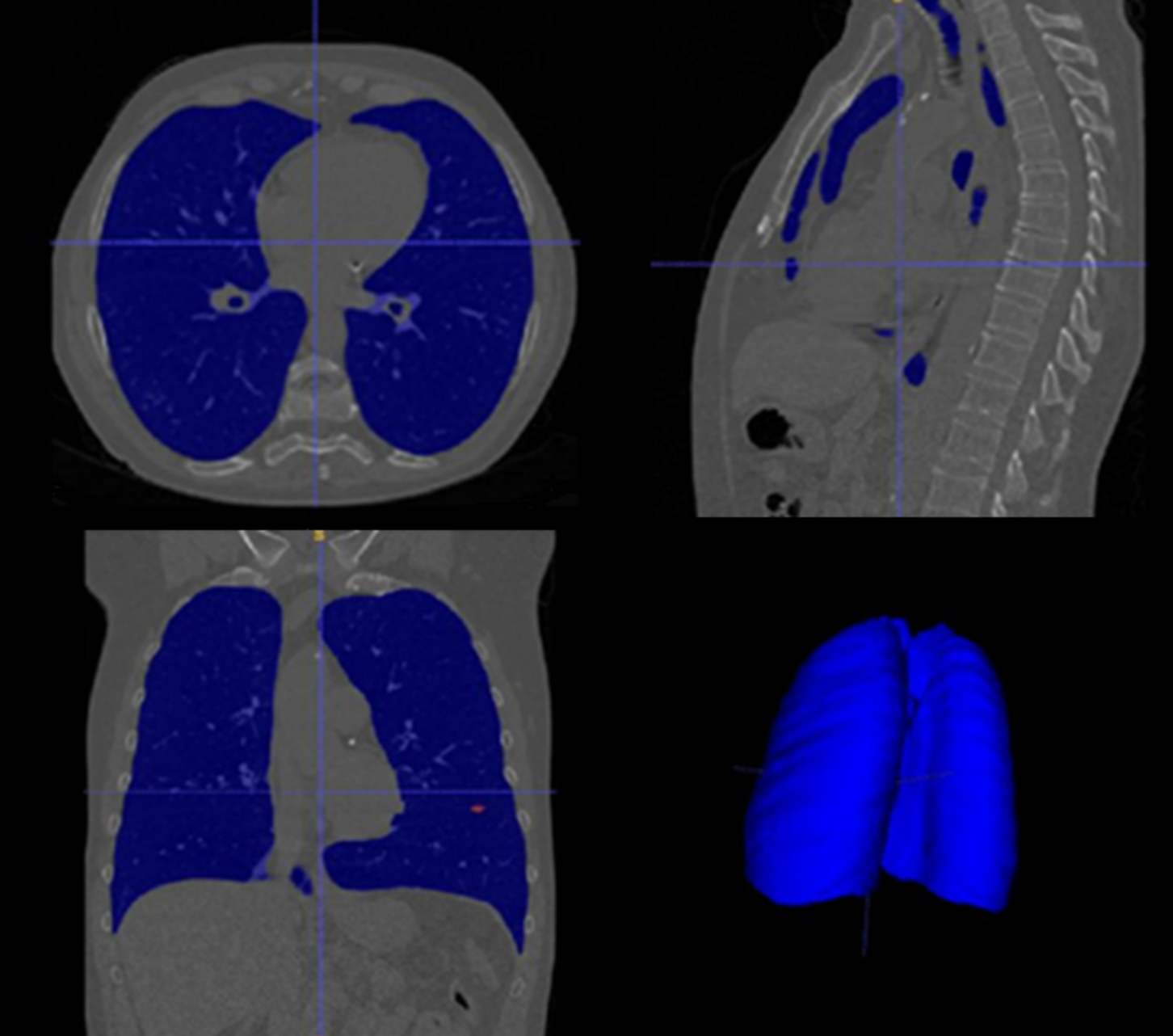
Patient_2 Normal

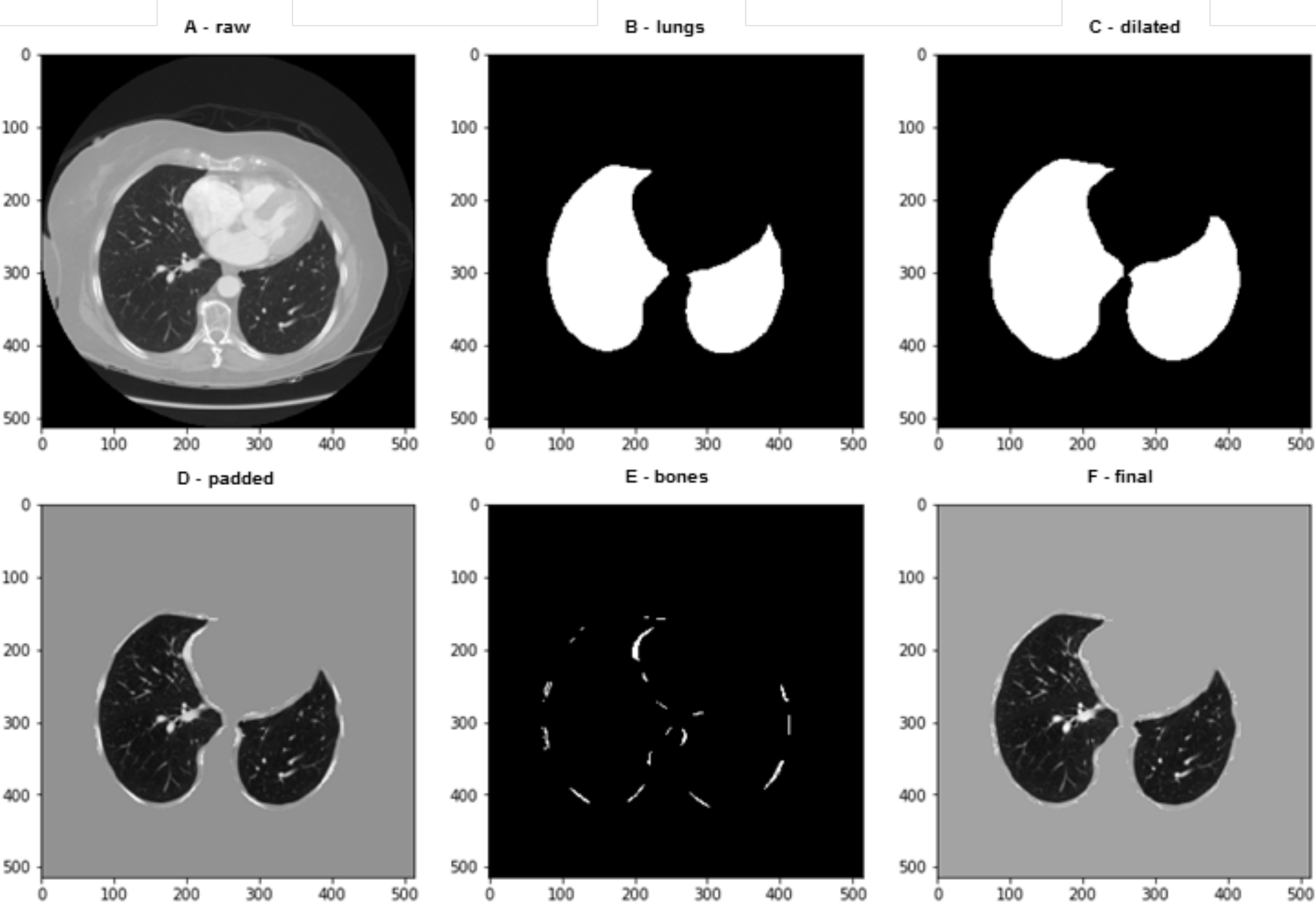
Pre-processing

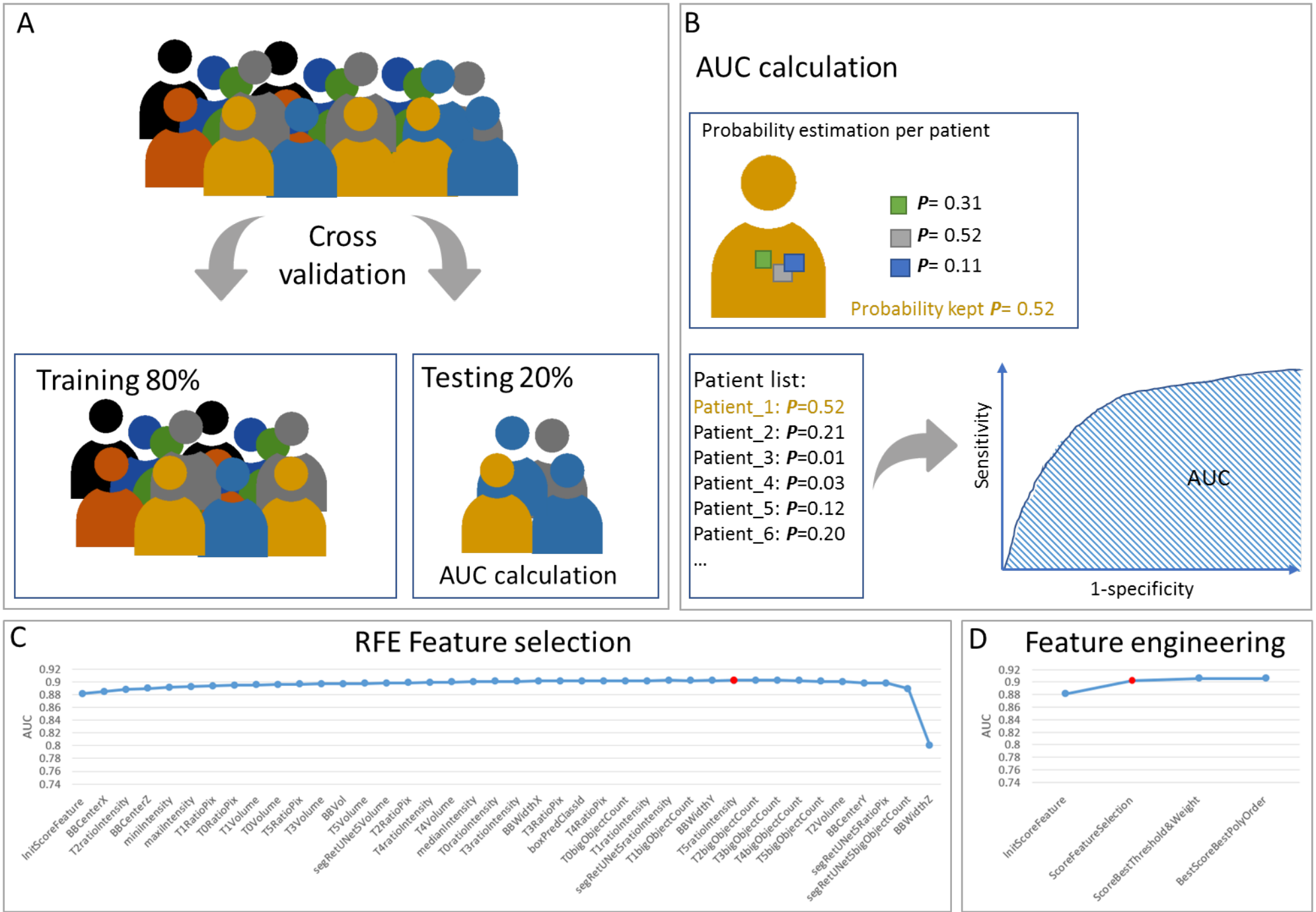
Nodule detection

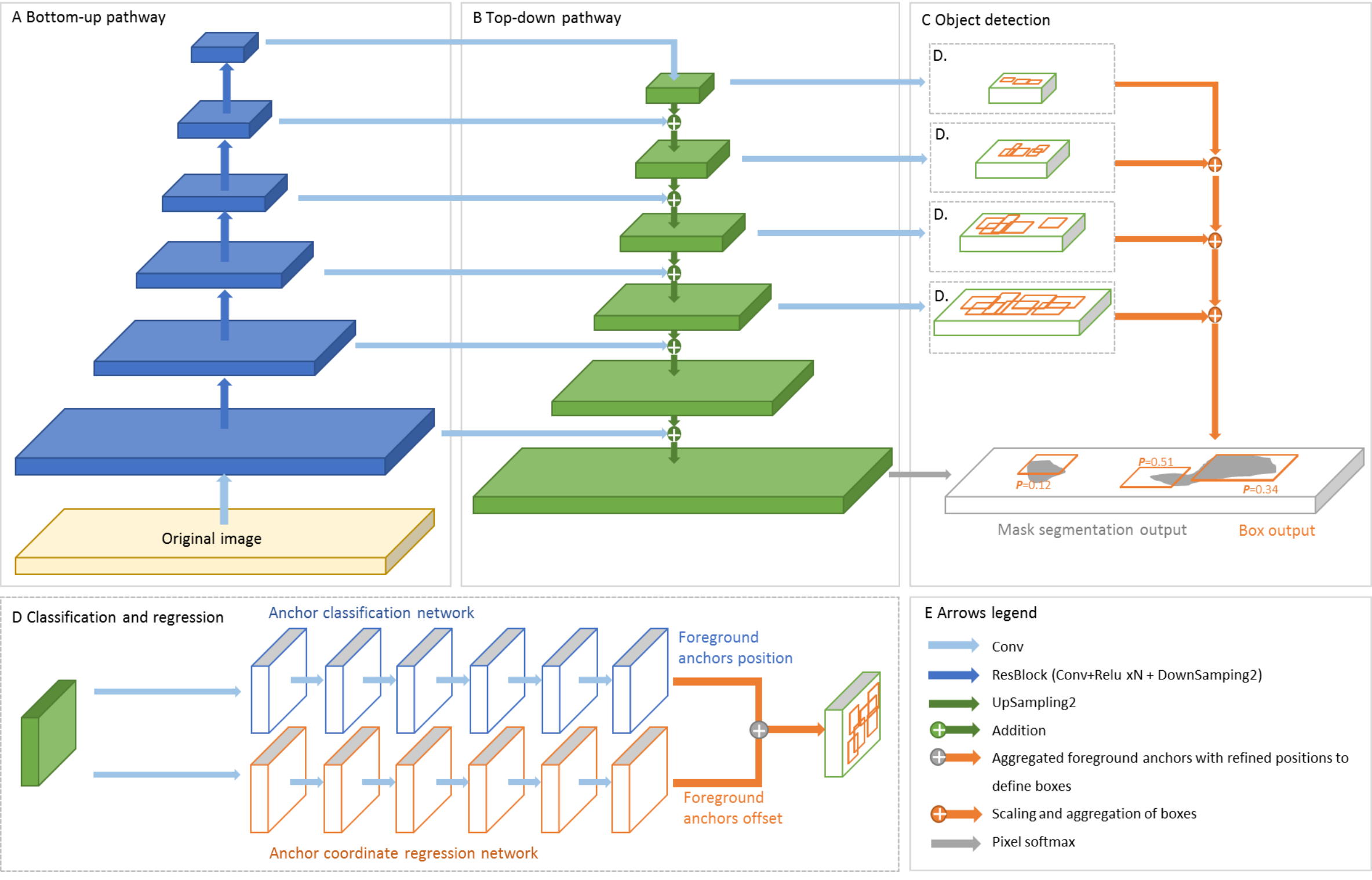
Nodule classification


0
100
200
300
400
500
0
100
200
300
400
500
0
100
200
300
400
500
0
100
200
300
400
500




A - raw
B - lungs
C - dilated
D - padded
E - bones
F - final
0
100
200
300
400
500


A
Cross validation
Training 80%
Testing 20%
AUC calculation
B
AUC calculation
Probability estimation per patient
P= 0.31
P= 0.52
P= 0.11
Probability kept P= 0.52
Patient list:
Patient_1: P=0.52
Patient_2: P=0.21
Patient_3: P=0.01
Patient_4: P=0.03
Patient_5: P=0.12
Patient_6: P=0.20
...
Sensitivity
AUC
1-specificity
C
RFE Feature selection
AUC
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
InitScoreFeature
BBCenterX
T2ratioIntensity
BBCenterZ
miniIntensity
maxIntensity
T1RatioPix
T0RatioPix
T1Volume
T0Volume
T5RatioPix
T3Volume
BBVol
T5Volume
segRetUNet5Volume
T2RatioPix
T4ratioIntensity
T4Volume
medianIntensity
T0ratioIntensity
T3ratioIntensity
BBWidthX
T3RatioPix
boxPredClassId
T4RatioPix
T0bigObjectCount
T1ratioIntensity
segRetUNet5ratioIntensity
T1bigObjectCount
BBWidthY
T5ratioIntensity
T2bigObjectCount
T3bigObjectCount
T4bigObjectCount
T5bigObjectCount
T2Volume
BBCenterY
segRetUNet5RatioPix
segRetUNet5bigObjectCount
BBWidthZ
D
Feature engineering
AUC
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
InitScoreFeature
ScoreFeatureSelection
ScoreBestThreshold&Weight
BestScoreBestPolyOrder


A Bottom-up pathway
B Top-down pathway
C Object detection
D.
D.
D.
D.
Original image
P=0.12
P=0.51
P=0.34
Mask segmentation output
Box output
D Classification and regression
Anchor classification network
Foreground anchors position
Foreground anchors offset
Anchor coordinate regression network
E Arrows legend
Conv
ResBlock (Conv+Relu xN + DownSamping2)
UpSampling2
Addition
Aggregated foreground anchors with refined positions to define boxes
Scaling and aggregation of boxes
Pixel softmax