

1 **Large scale validation of the M5L lung CAD on heterogeneous CT datasets.**

2 E. Lopez Torres,^{1,2, a)} E. Fiorina,^{3,2} F. Pennazio,^{3,2} C. Peroni,^{3,2} M. Saletta,² N.
3 Camarlinghi,^{4,5} M.E. Fantacci,^{4,5} and P. Cerello^{2, b)}

4 ¹⁾*CEADEN, Havana, Cuba*

5 ²⁾*INFN, Sezione di Torino, Torino, Italy*

6 ³⁾*Department of Physics, University of Torino, Torino,*
7 *Italy*

8 ⁴⁾*Department of Physics, University of Pisa, Pisa, Italy*

9 ⁵⁾*INFN, Sezione di Pisa, Pisa, Italy*

10 (Dated: 13 January 2015)

11 **Purpose:** M5L, a fully automated Computer-Aided Detection (CAD) system for
12 the detection and segmentation of lung nodules in thoracic Computed Tomography
13 (CT), is presented and validated on several image datasets.

14 **Methods:** M5L is the combination of two independent sub-systems, based on the
15 *Channeler Ant Model* as a segmentation tool (lungCAM) and on the Voxel Based
16 Neural Approach (VBNA). The lungCAM was upgraded with a scan equalisation
17 module and a new procedure to recover the nodules connected to other lung struc-
18 tures; its classification module, which makes use of a Feed-Forward Neural Network,
19 is based of a small number of features (13), so as to minimise the risk of lacking
20 generalisation, which could be possible given the large difference between the size of
21 the training and testing datasets, which contain 94 and 1019 CTs, respectively. The
22 lungCAM (standalone) and M5L (combined) performance was extensively tested on
23 1043 CT scans from 3 independent datasets, including a detailed analysis of the full
24 LIDC/IDRI database, which is not yet found in literature.

25 **Results:** The lungCAM and M5L performance is consistent across the databases,
26 with a sensitivity of about 70% and 80%, respectively, at 8 False Positive findings
27 per scan, despite the variable annotation criteria and acquisition and reconstruction
28 conditions. A reduced sensitivity is found for subtle nodules and Ground Glass Opac-
29 ities (GGO) structures. A comparison with other CAD systems is also presented.

30 **Conclusions:** The M5L performance on a large and heterogeneous dataset is stable
31 and satisfactory, although the development of a dedicated module for GGOs detec-
32 tion could further improve it, as well as an iterative optimisation of the training
33 procedure. The main aim of the present study was accomplished: M5L results do
34 not deteriorate when increasing the dataset size, making it a candidate for supporting
35 radiologists on large scale screenings and clinical programs.

36 **Keywords:** Lung CT; Computer Aided Detection (CAD); Image Processing; 3-D
37 segmentation; LIDC IDRI; ANODE09; Screening

a)Ernesto.Lopez.Torres@cern.ch

b)cerello@to.infn.it

38 I. INTRODUCTION

39 Lung cancer accounts for about 19% and 28% of cancer-related deaths in Europe¹ and the
40 United States of America², respectively. An improved prognosis would likely save thousands
41 of lives every year, with a very relevant impact on global health statistics. As it happened
42 for other types of cancer (e.g., breast cancer), an early diagnosis is expected to help in
43 optimising the effectiveness of treatment, improving its outcome and reducing the mortality.

44 Since lung cancer is most frequently detectable as non-calcified pulmonary nodules, Com-
45 puted Tomography (CT) is the most appropriate imaging modality for its early detection³.
46 The concept of screening, already adopted for breast cancer, is being considered for lung
47 cancer as well: several pilot programs based on low-dose high resolution CT were operated
48 worldwide⁴⁻⁶ during the last decade. Recent results reported by the National Lung Screening
49 Trial (NLST)⁷ show a statistically significant reduction (about 20%) of the 5-year mortality
50 in the branch subject to CT screening as compared to chest X-ray. The design and opera-
51 tion of large scale lung cancer screening programs is now being considered, with the goal of
52 maximising their effectiveness and minimising their cost.

53 Among the relevant issues to be addressed are:

- 54 • the optimisation of the nodule detection performance, in terms of sensitivity and
55 specificity, which could be based on double-reading. It was indeed observed that
56 a relevant fraction of lung nodules (20 to 35%) are missed in single-reader screening
57 diagnoses⁸. Moreover, the radiologist performance is strongly dependent on experience
58 and physical conditions such as stress and fatigue, which cause fluctuations in the inter-
59 and intra-radiologist sensitivity, respectively.
- 60 • the size of imaging data that must be coherently handled, since multi-detector helical
61 CT with thin collimation generates up to 600 two-dimensional images per scan.
- 62 • the amount of human resources (i.e., the number of radiologists) involved in the an-
63 notation process, since a careful reading of a high-resolution CT requires an average
64 time of several minutes.

65 In such a scenario, Computer-Aided Detection (CAD) algorithms could support radiol-
66 ogists with an automated identification and segmentation of small nodules, a signature of
67 possible early-stage disease. Several studies⁹⁻¹¹ reported an improvement in the sensitivity

68 of radiologists when assisted by CAD systems, in addition to a relevant time saving. Other
69 studies^{12,13} observe that the increase in detection rate is associated to an increase in the
70 number of false-positive findings. However, CAD systems act as detection rates equalisers
71 between observers of different levels of experience¹².

72 In order to be effectively introduced in the report-generating process, CAD systems must
73 provide an adequate performance (high sensitivity and as low as possible rate of false posi-
74 tives), properly validated on as large as possible a sample of CTs, so as to keep under control
75 the main sources of performance variability and degradation:

- 76 • the inter-center variability of acquisition setups, which turns into different properties
77 for the images (one above all: the reconstruction-related equivalent thickness of the
78 2D slices);
- 79 • the different annotation criteria adopted by different screening programs and/or by
80 different sites;
- 81 • the definition of a training sample that is representative of the features of the entire
82 population of the structures being searched for;

83 While the annotation-related variability can - to some extent - be parametrised with
84 a proper algorithm configuration (e.g., selection of findings with a radius larger than a
85 protocol-related minimum value), the slice thickness and the lack of generalisation issues
86 are related to two main conditions: the size and heterogeneity of the training, testing and
87 validation samples and the algorithm design.

88 With the goal of providing an adequate overall performance averaged over different types
89 of nodules, given the relatively small training sample size a small number of key features was
90 selected. However, should a CAD system be optimised for a specific category of nodules,
91 like ground glass opacities (GGO), a training sample larger than the presently available one
92 would be required.

93 The paper aims at validating the M5L CAD, which combines the lungCAM and VBNA
94 subsystems **and includes segmentation, nodule hunting and classification**, on the largest and
95 most heterogenous dataset available, so as to evaluate its readiness for application as a
96 support for screening programs and clinical practice.

97 The upgrade of the lungCAM sub-system with respect to the CAM segmentation al-
98 gorithm is discussed in detail; the VBNA sub-system, on the other hand, was already
99 described^{14,15}.

100 II. MATERIALS

101 Among the required features of a system for clinical and screening applications is the
102 capability to provide a performance independent of the dataset source: for that reason,
103 several datasets were analysed, collected both from screening programs and from clinical
104 practice.

105 A. LIDC/IDRI

106 The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative
107 (IDRI) provide the largest publicly available collection of annotated CTs¹⁶: 1018 CT scans
108 are available since 2011. LIDC/IDRI is a multi-center and multi-manufacturer database,
109 with CTs taken at different collimation, voltage, tube current and reconstructed slice thick-
110 ness. It provides a sample likely to realistically represent the input from a large scale
111 multi-center screening program as well as clinical practice. In order to capture the inter-
112 reader variability LIDC/IDRI provides, for each CT scan, four annotations made by different
113 expert radiologists, obtained with a two phase reading modality.

114 B. ITALUNG-CT

115 The ITALUNG-CT study⁵, carried on in Italy over the last decade, aimed at verifying
116 the effectiveness of screening in reducing the lung cancer mortality rate. A sample of 20
117 low-dose high-resolution CTs, acquired in the so-called screening setting (140 kV,70-80 mA,
118 1.25 *mm* reconstructed slice thickness) was made available for the validation. The scans
119 were annotated by 2 experienced radiologists, who were requested to identify nodules by
120 defining a cue point and a radius and by labelling the finding as relevant (diameter larger
121 than 5 *mm*, type 1) or not (diameter in the 3 – 5 *mm* range, type 2, a possible recent cancer
122 formation to be kept under control in follow-up sessions).

123 C. ANODE09

124 The ANODE09¹⁷ data set consists of 55 anonymised CT scans provided by the Utrecht
125 University Medical Center and originates from the NELSON study, the largest lung cancer
126 screening trial in Europe. 5 CT scans are made available together with the radiologist
127 annotations and can be used for training a CAD system; 50 scans can only be used for a
128 blind validation. Most of the database was randomly selected; however some CTs with a
129 large number of nodules were deliberately included. The ANODE09 annotation protocol
130 foresees the labelling of relevant nodules when their diameter is larger than 4 *mm*.

131 D. Training Dataset

132 The M5L CAD neural-network classifiers, for lungCAM and VBNA, were trained on 69
133 lung CT scans from LIDC/IDRI, 5 from ANODE09 and 20 from ITALUNG-CT, as discussed
134 in¹⁸. The 69 LIDC/IDRI CTs had already been used for training M5L to submit the results
135 to the ANODE09 challenge. For the full LIDC/IDRI dataset analysis, we decided to keep the
136 same training dataset so as to make the results directly comparable across the 3 databases.
137 Since one of the main purposes of this validation is to show that even without changing pa-
138 rameters the system performance is satisfactory, the algorithm parameters were not changed
139 and were the same for all 3 datasets for the present analysis.

140 The results were obtained on 949 CT scans from the LIDC/IDRI database (excluding
141 the 69 CT scans randomly selected from the training subset) 50 scans from ANODE09 and
142 20 from ITALUNG-CT.

143 For further reference the *shared lists* corresponding to each LIDC/IDRI subset have been
144 saved in The Cancer Imaging Archive online database by the National Cancer Institute and
145 are available for download at <http://cancerimagingarchive.net> with the following names:

- 146 • **LIDC_training_lungCAM**: 69 cases for training;
- 147 • **LIDC_test_lungCAM**: 949 cases for the validation.

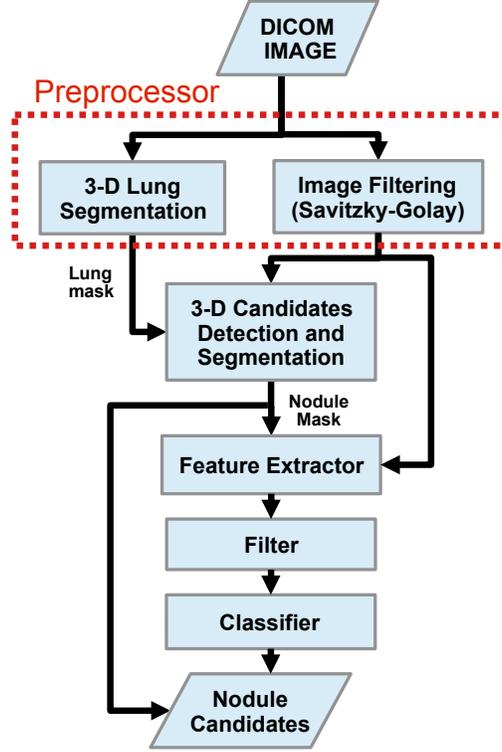


FIG. 1. lungCAM Block Diagram

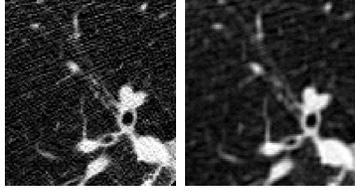
148 III. THE lungCAM ALGORITHM

149 The lung Channeler Ant Model (lungCAM) was developed by the MAGIC-5 Project¹⁹ as
 150 part of a multi-thread Computer Aided Detection (CAD) system for radiologist support in
 151 the lung cancer diagnosis, that also includes algorithms based on Region Growing (RGVP)²⁰
 152 (not supported anymore) and Voxel-Based Neural Analysis (VBNA)^{14,15}.

153 At the highest abstraction level, the lungCAM structure is a standard approach, as shown
 154 by the algorithm block diagram (Fig. 1): the preprocessing stage (equalisation and lung
 155 volume segmentation) is followed by a search for Regions Of Interest (ROIs), an analytical
 156 filter and a neural classifier.

157 Before starting the actual analysis, CT scans in DICOM standard format are preprocessed
 158 to reduce the noise contribution: each 2D slice is analysed with a Savitzky-Golay filter^{21,22}
 159 that provides noise reduction without loss of resolution. Fig. 2 shows an example of a 2D
 160 slice before and after the filtering stage.

161 From then on, every step of the lungCAM algorithm, including the features evaluation,
 162 is intrinsically 3-dimensional.



(a) (b)

FIG. 2. (a) Original slice image, (b) slice image after 2-D Savitsky-Golay filter

163 A. Lung Segmentation

164 The lung segmentation, described in detail elsewhere²³, proceeds according to four main
165 steps:

- 166 1. analysis of the CT Hounsfield Unit level distribution and evaluation of the intensity
167 threshold to be applied in the following stages;
- 168 2. 3D region growing of the lung volume with the detected threshold;
- 169 3. wavefront algorithm for the definition of the lung surface on the inner side and the
170 removal of the trachea and the main bronchi;
- 171 4. morphological closing with a cylinder from the outside in order to include pleural
172 nodules and close the holes left by vessels.

173 A check on the training/testing and validation data-sets confirmed that none of the
174 radiological findings were rejected at this stage.

175 B. Nodule segmentation

176 The segmentation algorithm is performed with the Channeler Ant Model (CAM)²⁴, based
177 on Virtual Ant Colonies and conceived for the segmentation of complex structures with
178 different shapes and intensity range in a noisy 3D environment. The CAM exploits the
179 natural capabilities of Virtual Ant Colonies to modify the environment and communicate
180 with each other by pheromone deposition. The ant life cycle is a sequence of atomic time
181 steps, during which the behaviour is determined by a set of rules that control the pheromone
182 release, the movements and the variations of the ant energy, a parameter related to breeding
183 and death.

184 The lung internal structures are segmented by iteratively deploying ant colonies in voxels
 185 with intensity above a pre-defined threshold (anthills). Ants live according to the model
 186 rules until the colony extinction: the pheromone deposition generates pheromone maps.

187 Each voxel visited by an ant during the life of a colony is removed from the allowed
 188 volume for future ant colonies. New ant colonies are iteratively deployed in unvisited voxels
 189 that meet the anthill requirement. By an iterative thresholding of pheromone maps a list of
 190 ROI candidates is obtained. ROIs with a radius larger than 10 *mm* are post-processed in
 191 order to disentangle nodules attached to internal lung structures like vessels and bronchi.

192 In order to speed up the ant deployment, the probability P_{ij} for a voxel to become the
 193 actual ant destination was changed from:

$$P_{ij}(v_i \rightarrow v_j) = \frac{W(\sigma_j)}{\sum_{n=1,26} W(\sigma_n)} \quad (1)$$

194 to:

$$P'_{ij}(v_i \rightarrow v_j) = 1 - P_{ij}(v_i \rightarrow v_j) + P_{ij}^{Int}(v_i \rightarrow v_j) \quad (2)$$

195 where $W(\sigma_j)$ depends on the amount of pheromone in voxel v_j and in $P_{ij}^{Int}(v_i \rightarrow v_j)$ is the
 196 same as $P_{ij}(v_i \rightarrow v_j)$ but substituting $W(\sigma_j)$ by the intensity of the voxel I_j . The resulting
 197 probabilities $P'_{ij}(v_i \rightarrow v_j)$ are normalised to a unitary total probability.

198 The new rule favours destination voxels with low integrated pheromone deposition and
 199 high Hounsfield Unit (HU) values, i.e. voxels with few visits: therefore, the colony expands
 200 faster in the 3D environment and the algorithm speed increases.

201 A limit to the maximum number of voxel visits $N_v(I_j)$ that a voxel j with HU intensity
 202 I_j receives from ants was also set to:

$$N_v(I_j) = N_{min} \left(1 + \frac{I_{max} - I_j}{I_{max}} \right) \quad (3)$$

203 where I_{max} is the maximum HU intensity value in the lung volume and N_{min} is a free
 204 parameter related to the algorithm speed, set to 5 for the present application.

205 Another limitation was related to the fact that for small low-intensity nodules the ant
 206 colony would extinguish too quickly to produce a pheromone image that could be identified
 207 by the threshold-based pheromone map analysis. The ant capability to explore low intensity
 208 voxels depends on the energy variation rate (4), i.e. on how many steps in low intensity

209 voxels ants can take on average before their energy drops to the death level. When objects
 210 are very small, also the initial random movement can play an important role in causing the
 211 premature colony extinction.

$$\varepsilon_{t+1}^k - \varepsilon_t^k = -\alpha \left(1 - \frac{\Delta_{ph}^k}{\langle \Delta_{ph} \rangle} \right) \quad (4)$$

212 The issue was addressed with a change in the ant colony evolution dynamics: the ant
 213 energy parameters (the initial energy ε_0 and its variation rate α) are initially set to values
 214 that cause a quick ant reproduction rate. Only when the colony population grows above 100
 215 units, the parameter values switch to the model default values, so as to avoid the exponential
 216 increase of the colony population: in such a way, a better pheromone image for small and
 217 low-intensity nodules is obtained without affecting the segmentation of large structures.

218 **1. Structure Segmentation**

219 The CAM is iteratively deployed in the right and left lungs, separately, as a segmentation
 220 method for the vessel tree and the nodule candidates. The first ant colony segments the
 221 vessel tree, starting from an anthill in the vicinity of its root. The segmented object is then
 222 removed from the original image and the coordinates of all its voxels are stored as a single
 223 Region Of Interest (ROI).

224 In the remaining image, iteratively, any voxel with intensity above a predefined threshold
 225 (-700 HU) is a new anthill and a colony deployed from there generates a pheromone image.
 226 When no more voxels meet the condition to become an anthill, the information provided by
 227 the global pheromone map is analyzed.

228 **2. Nodule Hunting**

229 The pheromone map analysis is also iterative: each voxel with a pheromone content above
 230 a minimum accepted value is used as a seed for a region growing with an adaptive threshold
 231 which is iteratively lowered until a minimum growth rate of the region is reached.

232 Every grown region with a radius in the 0.8 – 25 mm range is considered as a nodule
 233 candidate.

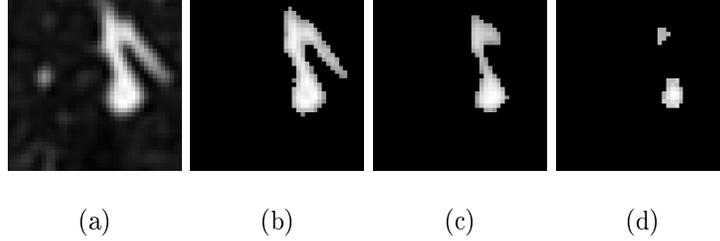


FIG. 3. Steps of separation procedure for juxta-vascular nodules. (a) Original image, (b) Pheromone segmentation, (c) Pass 1 with $R=1.5$ mm, (d) Pass 2 with $R=2.5$ mm. LIDC/IDRI case LIDC-IDRI-0039, nodule at 207,206,175

234 3. *Juxta-Vascular Nodules*

235 About 20% of relevant pulmonary nodules are segmented together with a vascular struc-
 236 ture they are connected to. If features were evaluated for the whole ROI, these nodules
 237 would typically be rejected by further filtering and classification.

238 In order to address the problem a dedicated algorithm module was developed. All the
 239 structures obtained from the pheromone map analysis with radius larger than 10 *mm* are
 240 further analysed in order to identify and disentangle spherical-like sub-structures. The
 241 10 *mm* value was empirically set based on the minimum size for attached structures that
 242 causes a relevant change in the ROI feature values.

243 Each voxel that belongs to the structure being analysed is averaged with the neighbours
 244 inside a sphere of radius R . Then, the average map is thresholded at the T_{ph} pheromone value
 245 again, resulting in a thinner object. Structures with a diameter smaller than R disappear
 246 (e.g., thin vessels attached to the nodules). However also the nodules shrink. In order to
 247 recover the nodule original size, the neighbours of each remaining voxel in the average inside
 248 a sphere of radius $R/2$ with value above $T_{ph} + T_{ph}/3$ in the original map are restored as part
 249 of the structure.

250 The procedure is repeated three times, with spheres of increasing radius ($R = 1.5, 2.5, 3.5$ *mm*)
 251 that generate sub-structures of increasing size. The output voxels of the three iterations are
 252 combined in logical OR to generate a final nodule candidate output mask, which is then
 253 treated as a ROI for further analysis.

254 Fig. 3 shows an example of separation of a juxta-vascular nodule from the vascular tree.

Geometrical features	Intensity-related Features
center of gravity $X_{i=x,y,z}$ (*)	Average
Radius (mm)	Average outside mask
Sphericity	Std. Deviation
Skewness of distance from X_i	Std. Deviation outside mask
Kurtosis of distance from X_i	Maximum
Volume (mm ³) (*)	Entropy
Fraction of voxel connected to pleura	Entropy outside mask

TABLE I. List of features extracted from the nodule output mask. Features labeled with the asterisk were not used in the classification stage.

255 C. ROI Features

256 The choice of a suitable set of ROI features is a key to the success of the filtering and
257 classification stages. Ideally, any computable quantity which is expected to show a different
258 pattern for true nodules and false candidates would be a useful feature. However, the use of
259 a large number of features on a small training dataset could bias the classifier and cause a
260 loss of generality.

261 The choice to select a small number of features for the neural classifier training aims at
262 optimising the generality and keeping the performance stable as the validation dataset size
263 increases.

264 A set of 13 features was selected for the nodule candidate analysis, according to the
265 following criteria:

- 266 • 3D **spatial** features **which** are invariant to rotation and translation and can disen-
267 tangle spherical-like structures from ROIs originating from vessel parts or lung walls;
- 268 • features based on the voxel HU intensity, so as to capture density patterns;
- 269 • the fraction of ROI voxels attached to the walls of the lung volume is crucial in
270 distinguishing internal and juxta-pleural nodules, which are characterised by a different
271 shape; therefore, its use allows the classification of both the subsamples with the same
272 neural network.

273 The radius R is defined as the average distance of ROI voxels from the center of gravity
 274 times $4/3$, so as to be equal to the radius of the sphere if the ROI were perfectly spherical.

275 The center of gravity coordinates X_i are computed using the HU values as weights, with
 276 an extra weight-factor of 0.1 for the voxels on the ROI surface:

$$X_i = \frac{\sum_{k=1}^N I_k r_{k,i} \varepsilon_k}{\sum_{k=1}^N I_k \varepsilon_k}, \varepsilon_k = \begin{cases} 1 & k \in \text{inside voxel} \\ 0.1 & k \in \text{surface voxel} \end{cases} \quad (5)$$

277 where the k index runs over all the voxels in the ROI, I_k is the intensity associated with
 278 the voxel k in HU units, i is equal to 1,2 and 3 for x, y and z , respectively, and $r_{k,i}$ is the
 279 position vector of voxel k . The extra weight factor ε_k helps to better locate the center of
 280 gravity in case of ROIs that include some pieces of vessel or pleura surface, by suppressing
 281 the contributions of nodule substructures with a high surface to volume ratio.

282 The *Sphericity* is defined as the ratio of the ROI volume to the volume of a sphere with
 283 radius R equal to that of the ROI.

284 The *Fraction of voxel connected to the lung volume surface* is calculated by dividing
 285 the number of voxels connected to the lung mask by the number of surface voxels of the
 286 segmented object.

287 The features labeled with *outside mask* in Table I are computed by enlarging the original
 288 segmentation using a spheroidal structuring element of 1.5 mm of radius.

289 In the present work a further optimisation of the set of features was not carried on,
 290 although the size of the training and validation dataset would allow it: our goal is to
 291 demonstrate that, even with a training based on a fairly small number of lung nodules, a
 292 CAD system can be predictive and keep its performance on large validation datasets such
 293 as LIDC/IDRI.

294 D. Filtering

295 The average number of ROIs after the nodule hunting, depending on the number of slices,
 296 ranges between several hundreds to few thousands per CT scan, a number far too large to be
 297 used as input for a neural network classifier. The vast majority of findings is easily rejected
 298 with an analytical filter based on correlations between the radius, the sphericity and the
 299 fraction of voxels connected to the lung mask. Fig. 4 shows the correlation between the

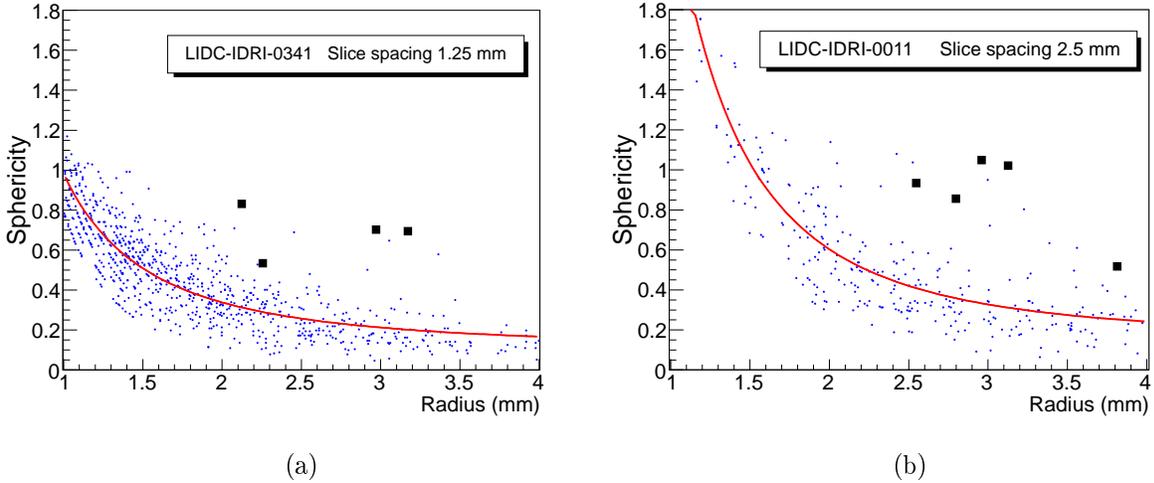


FIG. 4. Distribution of *Sphericity* as a function of *Radius* for two scans from LIDC/IDRI, with slice spacing of 1.25 mm (a) and 2.5 mm (b). Black square markers are used for the candidates corresponding to true nodules.

300 *Sphericity* and the *Radius* of nodule candidates, with the true nodules highlighted as black
 301 squares: it is clear that the correlation can be used to filter most of the false positive (FP)
 302 findings.

303 However, the discrete nature of the CT images implies that geometrical features depend
 304 to some extent on the voxel size, particularly for small ROIs with few voxels. Some CAD
 305 systems²⁵ have adopted a down-sampling approach, so as to obtain a comparable slice-
 306 spacing in all the dataset. In LIDC/IDRI, the CT slice spacing ranges from 0.6 mm to
 307 3.0 mm: as a consequence, the distribution of values for some features, like the *Sphericity*,
 308 shows different values depending on the slice spacing.

309 Fig. 4 (a) and (b) show the correlation between *Sphericity* and *Radius* values for nodule
 310 candidates obtained from CTs with a slice spacing of 1.25 mm and 2.5 mm, respectively:
 311 the correlation depends on the slice spacing and the *Sphericity* shifts to higher values in the
 312 2.5 mm case. It is therefore not possible to use the same filter function on the whole dataset
 313 without compensating for this effect.

314 The correlation between the ROI *Sphericity* and *Radius* was then equalised by fitting it
 315 for each single CT with the $S = a/R^b + c$ function, represented by the red line in Fig. 4.

316 The equalised global distribution is then obtained, for all the CT scans, as the difference
 317 between the original *Sphericity* of a nodule candidate and the threshold *Sphericity* obtained

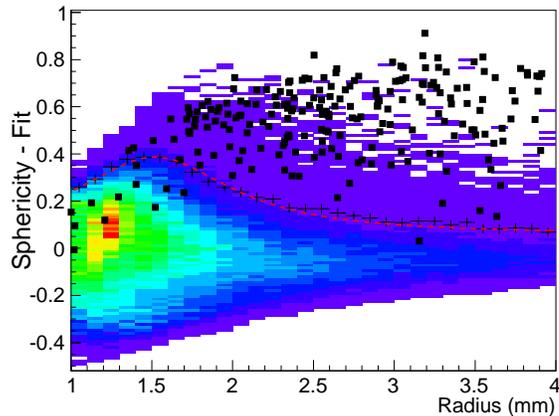


FIG. 5. Merged plot of the difference between *Sphericity* and *Sphericity-Fit* as a function of the *Radius* for all training scans. The black crosses are 2 standard deviation above the average for each bin and were used as reference to fit the final filter function represented by the red line.

318 from the single CT fit (fig. 5). For each bin along the x axis, the black crosses correspond
 319 to the average value plus 2 standard deviations of the sphericity difference and were used as
 320 reference points to fit the final filter function, represented by the red dashed line. All the
 321 ROIs with sphericity differences smaller than the filter function value were then discarded.

322 In addition to the sphericity-related selection, two other filtering conditions were applied
 323 to the nodule candidates: the *Fraction of voxels connected to lung surface* is required to be
 324 less than 0.6 and the *Radius* must be larger than 1.2 mm.

325 Irregular structures are filtered with these criteria. The CT equalisation and filtering
 326 procedure dramatically reduces the average number of FP findings per scan, from about
 327 1000 to about 50, a value which is appropriate as input for training and running a neural
 328 classifier.

329 The filtering process also reduces the pre-classification sensitivity to about 75 – 90%,
 330 depending on the input dataset.

331 E. The Neural Classifier

332 A feed forward neural network (FFNN) was selected as nodule candidate classification
 333 method. The training sample was made of 20, 5 and 69 CTs from the ITALUNG-CT,
 334 ANODE09 and LIDC/IDRI databases, respectively, for a total of 216 relevant nodules. The

335 training was carried on in cross-validation mode. The FFNN configuration was defined as
336 follows: 13 input neurons, 1 hidden layer with 25 neurons and 1 neuron in the output layer,
337 representing the probability of the finding to be relevant.

338 The choice implies that the overall performance is not fully optimised, since it aims at
339 proving the algorithm generality in realistic training conditions (i.e., a training sample much
340 smaller than the validation one. In view of a future application of the lungCAM CAD in
341 screening programs or clinical practice, the optimisation can be achieved by iteratively using
342 training samples of increasing size. Furthermore, demonstrating a generalisation capability
343 is, at the present development stage, even more important than optimising the sensitivity
344 on a selected dataset.

345 IV. RESULTS

346 M5L was validated on the datasets described in section II. The observed performance is
347 fairly stable, both as a function of the dataset and as a function of the number of scans (for
348 LIDC/IDRI).

349 Since the improvements described in this paper are related to the lungCAM sub-system
350 upgrade, results for lungCAM alone are discussed in detail, while the M5L performance is
351 presented on the overall dataset.

352 A. Labelling and Score

353 The labelling rule proposed in²⁵ was adopted: the *gold standard* information for the
354 ITALUNG-CT and ANODE09 includes cue points near the center of each nodule and its
355 radius, while for the LIDC/IDRI dataset the available manual segmentation is used. A
356 candidate is labeled a TP if its segmentation includes a true cue, a FP otherwise.

357 Since cue points for the LIDC/IDRI nodules are not provided, a different labelling method
358 was adopted: a CAD finding is considered a TP if the centroid of the segmented ROI is
359 contained within (any of) the radiologist segmentation(s).

360 The *Gold Standard* reference for the LIDC/IDRI dataset used for training the FFNN was
361 defined as the group of nodules with diameters $>3\text{ mm}$ annotated by at least 2 radiologists.
362 This definition is the closest to the annotation protocols of the ITALUNG-CT and ANODE

Database	Scan	True nodules	TP	FP/CT	Sensitivity (%)
ITALUNG-CT	20	39	32	38.6	82
ANODE09 test	5	39	30	16.5	76.9
LIDC training	69	138 (at least 2 rad.)	123	38.5	89.1
LIDC test	949	1747 (at least 2 rad.)	1421	52.1	81.3
TOTAL	1043	1943	1606	50.7	82.6
ANODE09	50	207	(149)*	(18.7)*	(72.4)*

TABLE II. lungCAM nodule detection performance after filtering. Numbers marked with (*) in the ANODE09 dataset were estimated from the FROC curve.

363 datasets, both based on double-reading.

364 Nodules with a diameter >3 mm that were annotated by 1 radiologist and *non-nodules*
365 were considered as "not-relevant" structures, i.e. neither true nor false findings, and were
366 ignored in the evaluation of the FROC curves.

367 B. Nodule Detection and Segmentation Performance

368 Table II shows the sensitivity of the lungCAM segmentation stage, after filtering, for each
369 database.

370 The ANODE09 (50) sample is excluded, since information about true nodules is not
371 publicly available.

372 The lungCAM performance is quite stable on the different databases, within the statistical
373 error, even though the image parameters, the acquisition and reconstruction conditions and
374 the annotation protocol are heterogeneous. This feature is particularly important in case of
375 both screening and clinical environment, with imaging studies coming from many sources.

376 Table III shows the average Jaccard index computed for 3 different consensus levels by
377 the radiologists, expressed in terms of probability map (PMAP) levels. A PMAP level,
378 associated to each voxel belonging to a nodule, is defined as the ratio between the number
379 of radiologists that included the voxel in the nodule and the total number of radiologists
380 that performed the annotation.

381 Messay *et al*²⁵ declare an average value of about 63% for 68 nodules reported by three

consensus / PMAP level	average Jaccard index
at least 2 / $\geq 50\%$	0.50 ± 0.20
at least 3 / $\geq 75\%$	0.57 ± 0.19
4 / = 100%	0.60 ± 0.18

TABLE III. Average Jaccard Index between the lungCAM nodule candidates and the manual segmentations.

382 radiologists at PMAP $>50\%$.

383 The nodule hunting sensitivity as a function of the nodule size is quite stable between 80%
384 and 90%, with the exception of small and large nodules, for which it drops to about 70%.
385 The actual nodule size for LIDC/IDRI is obtained by taking the largest *Radius*, obtained
386 from the radiologist contour/segmentation of each nodule.

387 C. lungCAM Performance

388 The performance is evaluated in terms of Free-response Receiver Operating Characteristic
389 (FROC) curves. The *LIDC test* and the *ANODE09* databases, not used at all in the training
390 or optimisation processes, provide a large and heterogeneous validation dataset.

391 The trained FFNN was applied to 949 LIDC/IDRI and 50 ANODE09 scans and the
392 lungCAM performance (Fig. 6) is very similar despite the differences between the datasets.

393 Since the LIDC/IDRI is a very heterogeneous database and only 69 out of 1018 were
394 used scans for training (and similarly 5 out of 55 for *ANODE09*), the results indicate a
395 satisfactory generalisation capability of the lungCAM system. The *ITALUNG-CT* FROC,
396 obtained with a training set that only included nodules from the 69 *LIDC training* and the
397 5 *ANODE09* CT scans, is also compatible.

398 A full statistical comparison across the three datasets is difficult, mostly because of the
399 different requirements in the annotation protocol: LIDC/IDRI, ANODE09 and *ITALUNG-*
400 *CT* foresee a nodule cutoff at a diameter of 3, 4, 5 mm, respectively, introducing a systematic
401 difference which can't be neglected and is hard to evaluate with the available information.
402 When possible, the statistical uncertainty was evaluated in terms of confidence interval: both
403 in the relevant working range ($6 < \text{FP}/\text{scan} < 8$) and in the sensitivity rising edge (FP/scan
404 < 2), the statistical uncertainty on FP/scan for ANODE09 and *ITALUNG-CT* dominates,

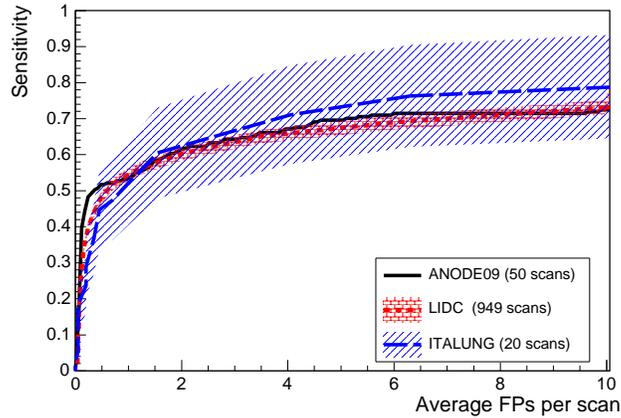


FIG. 6. lungCAM FROC curves for the *LIDC test* and the *ANODE09* validation datasets. The *ITALUNG-CT* FROC, obtained excluding its 20 scans from the training dataset, is also included for reference. The error bands show the statistical error on the LIDC/IDRI and *ITALUNG-CT* sensitivity.

405 given the relatively low number of true findings, so the results are fully compatible with
 406 LIDC/IDRI.

407 The error bands, showing the statistical uncertainty for LIDC/IDRI and *ITALUNG-CT*,
 408 confirm the compatibility of the results, although at least two large-size datasets would be
 409 required for a more stringent verification. The ANODE09 error band can't be evaluated as
 410 the number of true findings is unknown.

411 1. *LIDC/IDRI*

412 In order to validate the approach based on the equalisation of dimensional parameters
 413 (expressed in *mm*) via the fitting procedure on FP findings before the filtering stage, the
 414 FROC curves for three ranges of slice spacing were computed separately and compared (Fig.
 415 7): the results show a compatibility within 5% over the full FP range, with a slightly better
 416 sensitivity for larger (smaller) slice spacing below (above) 2 FP/scan.

417 The LIDC/IDRI database provides the detailed nodule segmentation for nodules with a
 418 diameter >3 mm, as well as information on several features: radiologists ranked subjective
 419 characteristics of the nodules such as subtlety, internal structure, spiculation, lobulation,
 420 shape sphericity, solidity, margin and likelihood of malignancy.

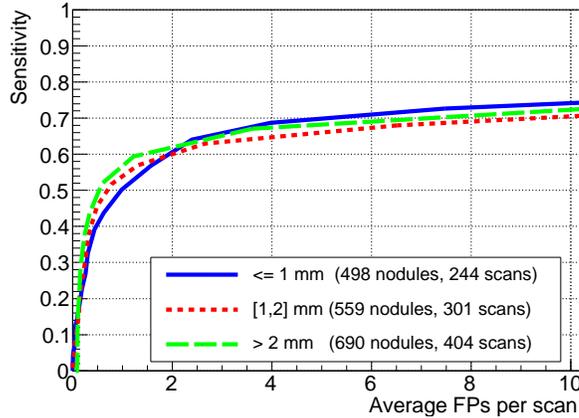


FIG. 7. FROC curves on *LIDC test* subsets corresponding to different scan slice spacing range (SR < 1 mm, 1 mm $< \text{SR} < 2$ mm, $\text{SR} > 2$ mm). The number of TP nodules and the number of scans for the different sub-samples are statistically comparable.

421 The availability of this classification allows the analysis of results as a function of the
 422 nodule type, helping in understanding the strengths and weaknesses of the lungCAM algo-
 423 rithm.

424 Fig. 8 shows the FROC curves for 3 groups out of 5 available malignancy rating: Unlike
 425 cancer (values of 1 and 2), Intermediate (3) and Highly Suspicious (4 and 5), represented
 426 by 526, 798 and 423 nodules, respectively. No information about malignancy in the training
 427 part was provided to the FFNN, so the classifier is expected to perform comparably for
 428 each type of lesion. The better performance at low FP values for unlike cancer nodules is
 429 probably related to the fact that they are typically calcified and therefore easier to detect.

430 Texture features are not directly included as input to the classifier; however, spherical-like
 431 objects are expected to be detected better than others. The analysis of non-solid lesions or
 432 ground glass opacities (GGO), which represent about 12% of the sample, shows they are
 433 poorly segmented and normally rejected during the filtering stage.

434 Another interesting characteristic that could affect the system performance is the subtlety
 435 score. Subtle nodules are hard to detect by both CADs and radiologists, since in general
 436 their HU intensity is very similar to the surrounding background and they are likely to be
 437 missed.

438 Subtle nodules represent about 7% of the *LIDC test* dataset and only about 20% of
 439 them are detected at 8 FP/scan. In general CADs need a fine tuning to improve the

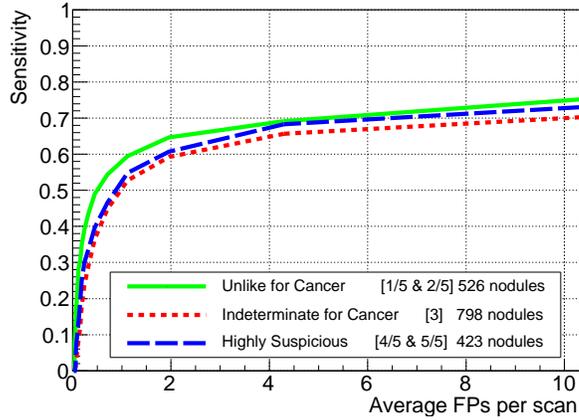


FIG. 8. FROC curves for different malignancy ratings on the *LIDC test* dataset. Malignancy scores were grouped as follows: Unlike cancer (1 and 2), Intermediate (3) and Highly Suspicious (4 and 5), corresponding to 526, 798 and 423 nodules, respectively.

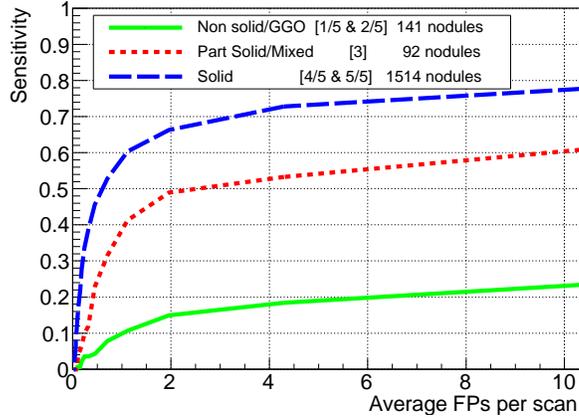


FIG. 9. FROC curves for different texture ratings on the *LIDC test* dataset. Texture score were grouped as follows: Non Solid/GGO (1 and 2), Part Solid/Mixed (3) and Solid (4 and 5), corresponding to 141, 92 and 1512 nodules respectively.

440 detection of subtle nodules without increasing significantly the FP rate. However, since
 441 also the radiologist sensitivity is likely to be smaller for subtle nodules, before starting any
 442 optimisation the CAD FP findings should be carefully analysed by radiologists, so as to
 443 identify possible subtle TP findings overlooked in the first round of annotation.

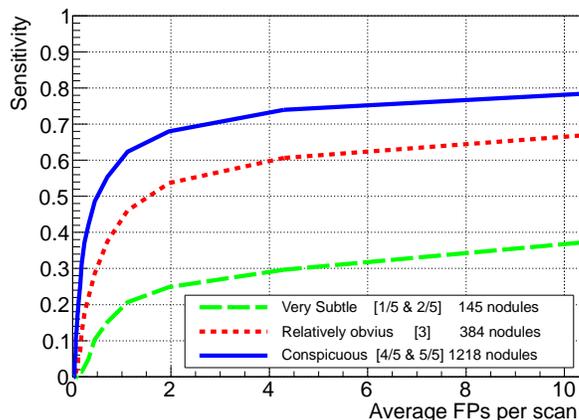


FIG. 10. FROC curves for different subtlety ratings on the *LIDC test* dataset.

444 2. ANODE09

445 The purpose of the ANODE09 challenge¹⁷ was to provide a database of CT scans from
 446 a lung cancer screening trial that would allow a fair blind evaluation of CAD algorithms
 447 under the same conditions and with the same metric. The only factor causing differences in
 448 results would then be the intrinsic CAD system performance, not the data or the details of
 449 the evaluation procedure.

450 ANODE09 results were scored with a metric that emphasises the performance at low FP
 451 values: the overall score of a system is calculated as the average of sensitivity values sampled
 452 at specificities 1/8, 1/4, 1/2, 1, 2, 4, and 8 FPs per scan.

453 The lungCAM sensitivity at these FP/scan values for all the *relevant* nodules and for
 454 subsets related to different nodule features are summarised in Table IV, while the FROC
 455 curves for each type of nodule are shown in Fig. 11. The early development version of the
 456 CAM algorithm, that joined the ANODE09 challenge in 2009 scored 0.254; the lungCAM
 457 as described in this work now scores 0.564, with a remarkable improvement.

458 D. Combining CADs

459 From the results of each system presented in the previous section, it is clear that ungCAM
 460 could be improved in the future, as any other CAD system, focusing on specific weaknesses.
 461 However, one quick and effective way to improve is to combine the results of different algo-
 462 rithms, as demonstrated in¹⁷ for the ANODE09 challenge participants.

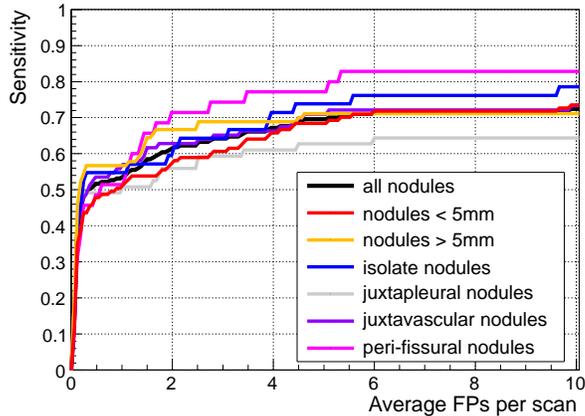


FIG. 11. FROC curves on the 50 CTs ANODE09 validation dataset for the different nodule types.

FPs/case	1/8	1/4	1/2	1	2	4	8	Average
Small nodules	0.359	0.436	0.478	0.513	0.572	0.658	0.718	0.533
Large nodules	0.478	0.555	0.566	0.566	0.666	0.689	0.711	0.605
Isolated nodules	0.428	0.524	0.547	0.547	0.595	0.714	0.762	0.588
Vascular nodules	0.430	0.488	0.535	0.570	0.628	0.663	0.721	0.576
Pleural nodules	0.440	0.491	0.491	0.508	0.559	0.610	0.644	0.535
Peri-fissural nodules	0.314	0.457	0.485	0.514	0.714	0.771	0.828	0.584
All nodules	0.410	0.488	0.517	0.536	0.613	0.671	0.715	0.564

TABLE IV. ANODE09 scoring: lungCAM sensitivity at the seven sampling points on the FROC curve and average score value.

463 Fig. 12 shows the results obtained when combining the lungCAM and VBNA M5L sub-
 464 systems on the 949 scans of the *LIDC test* dataset (949 scans).

465 The M5L sensitivity at 8 FP/scan reaches 80% which, given the size and heterogeneity
 466 of the dataset, is quite remarkable.

467 In the case of *ANODE09* the combined performance on the validation set, whose FROC
 468 is shown in Fig. 13, reaches a sensitivity score of 0.619. If the RGVP sub-system is added,
 469 M5L slightly outperforms IsiCAD with a score of 0.64. Further combinations provides even
 470 better results: M5L and IsiCAD score 0.752, which further improves to 0.760 when adding
 471 FlyesScan.

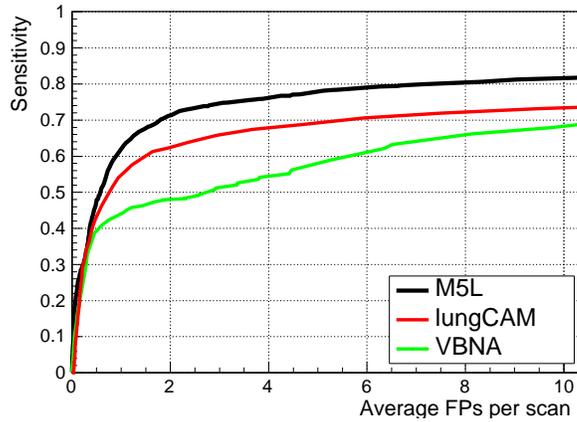


FIG. 12. FROC curves of the lungCAM and VBNA M5L sub-systems and their combination on the *LIDC test* validation dataset (949 scans).

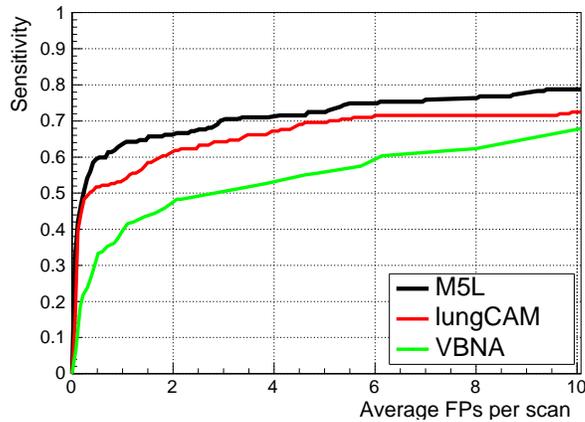


FIG. 13. FROC curves of the lungCAM and VBNA M5L sub-systems and their combination on the *ANODE09* validation dataset (50 scans).

472 **V. DISCUSSION**

473 CAD systems developed by academic research groups were reviewed in various papers^{26,27}:
 474 it is extremely difficult to make a fair comparison between all these CAD systems, mainly
 475 because of the difference in the definition of the properties of training, testing and validation
 476 datasets, the use of private datasets, insufficient statistics and sometimes different perfor-
 477 mance evaluation metrics. Some of them²⁸⁻³¹ have analysed LIDC/IDRI subsets, but none
 478 was tested on the full dataset.

479 Recently, two papers based on large validation datasets were published. One of them

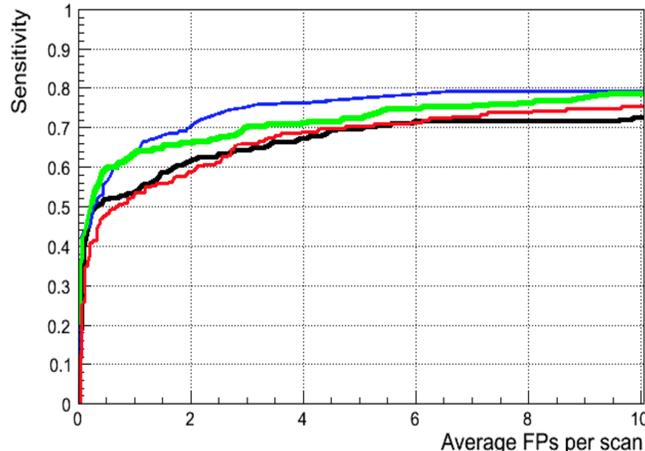


FIG. 14. FROC curves for IsiCAD (blue), FlyerScan (red), lungCAM (black) and M5L (green) on the 50 *ANODE09* validation scans.

480 discusses the performance of a CAD system in the NELSON⁴ screening trial³². A direct
 481 full comparison with the results presented in this work is not possible, as we do not have
 482 access to the full NELSON dataset; however, the results will be compared to a subsample of
 483 data from the NELSON screening program, made available through the ANODE09 study¹⁷.
 484 The other paper²⁵ presents results on ANODE09 and on the Lung Image Database Con-
 485 sortium (LIDC/IDRI)¹⁶ database, the largest available public database. This represent the
 486 opportunity to compare performance in almost the same conditions.

487 Commercially developed systems are usually bound to a specific hardware, so they are
 488 tuned for specific acquisition and reconstruction conditions. Besides, available algorithms
 489 are certified for reporting nodules above 4 *mm* in diameter^{33,34} and usually have a fixed
 490 threshold, so the comparison on the same dataset is not possible.

491 **In the following discussion, only published data are considered for a comparison with our**
 492 **results.**

493 The easiest comparison of lungCAM and M5L is with other participants in the AN-
 494 ODE09 study. The FROC curves for IsiCAD³², FlyerScan²⁵, lungCAM and M5L on the 50
 495 ANODE09 validation scans are presented in Fig. 14. The overall sensitivity score for the
 496 four systems is 0.632 (IsiCAD), 0.552 (FlyerScan), 0.564 (lungCAM) and 0.619 (M5L). Val-
 497 ues of sensitivity scores and FROCs curves on *ANODE09* were provided by the ANODE09
 498 challenge organisers (personal communication, 2014), as they are not publicly available on
 499 the ANODE09 website yet.

500 IsiCAD³² was developed at the University Medical Center Utrecht, the Netherlands, by
501 the group who organised the *ANODE09* study. It is based on shape index and curvedness
502 features and detects nodule candidates with a pre-classification sensitivity of about 97% at
503 about 700 FP/scan. The false-positive reduction consists of two consecutive classification
504 steps using k-Nearest-Neighbor (kNN) classifiers and the feature selection was carried out
505 by ‘Sequential Forward Floating Selection’. IsiCAD has the best performance, but it has
506 the advantage of having been trained over 722 scans from the NELSON screening program,
507 i.e. the same data source of the *ANODE09* validation dataset.

508 FlyerScan²⁵ was trained on 90 cases provided by the University of Texas Medical Branch.
509 It implements a simple and powerful combination of thresholding and opening operations to
510 segment the nodules candidates, which are detected at a 92.3% pre-classification sensitivity
511 at about 500 FP/scan (value provided for nodules >3 mm annotated by at least one radiol-
512 ogist on 84 LIDC/IDRI cases available in 2008). The algorithm was carefully optimised to
513 select the best features and the results of two classifiers using a different numbers of features
514 were compared. With the best classifier and using 40 features on the LIDC/IDRI 84 cases,
515 FlyerScan provides a sensitivity of 80.4% at 3 FP per scan. The same conditions were used
516 to analyse the *ANODE09* validation dataset, with the results shown in Fig. 14.

517 The overall performance of the lungCAM system on the ANODE09 and ITALUNG-CT is
518 comparable to other algorithms. The lungCAM is more selective in the filtering stage, with
519 an average post-filter sensitivity of about 80%, to be compared with about 90% reached by
520 other methods. However, the lungCAM classifier is then fed with a smaller number of false
521 positive findings and performs very well, bringing the overall sensitivity to the level of other
522 CAD systems.

523 Since no publication analysing the full LIDC/IDRI dataset (1018 scans) could be found,
524 the lungCAM (and M5L) results cannot be compared to other methods under the same
525 conditions.

526 Taking into account that the *Gold Standard* condition is defined as the consensus of at
527 least 2 radiologists out of 4 (i.e., the less restrictive possible condition), that the algorithm
528 was not optimised on the validation dataset, the M5L performance ($\sim 80\%$ at 6 FP/scan)
529 is considered satisfactory.

530 Some other CAD systems report sensitivities larger than 80%, but those results were
531 obtained on smaller datasets and should probably be confirmed in a configuration closer to

532 the actual clinical/screening operating conditions.

533 The M5L main limitation affecting the sensitivity lies in the segmentation and filtering
534 stage, where most of the findings, corresponding mainly to low intensity GGOs and subtle
535 nodules (9.5% and 5.5% of the sensitivity loss, respectively), are missed. The remaining
536 missing structures, accounting for about 4% of the sensitivity, are typically rejected in the
537 classification stage. The development of dedicated optimised modules for the segmentation
538 of GGOs and subtle structures is therefore the main task that could provide a significant
539 improvement. The optimisation of the neural network classifier is also likely to allow a slight
540 improvement on the sensitivity and a better rejection of the false positive findings, with a
541 shift to the left of the FROC. A larger training dataset could also be used to improve the
542 representation of every type of nodule and therefore the NN performance.

543 The sensitivity could also be further improved by extending the concept of subsystem
544 result combination to other algorithms, even developed by other research groups, as long as
545 their results are compliant with a fairly simple standard format for the CAD findings. and
546 integrating them in the final combination module.

547 However, a FP/scan value in the 4-8 range is commonly accepted by radiologists, as long
548 as a quick browsing of the CAD results is possible to minimise the FP rejection time. In
549 parallel, a clinical validation is planned, where the M5L impact on the radiologist perfor-
550 mance will be assessed: the gold standard obtained by the revision of the radiologist initial
551 annotation based on the M5L results will be compared to the initial result by the radiologist
552 and M5L alone.

553 VI. CONCLUSIONS

554 The M5L lungCAM sub-system includes two new modules, providing:

- 555 • the identification of nodules connected to internal lung structures;
- 556 • the equalization of CT scans, that allows the use of a common filtering function based
557 on the correlation between the candidate nodule *Sphericity* and *Radius*.

558 The above discussed results show that the M5L performance on a large and heteroge-
559 nous dataset is stable and satisfactory, although the development of a dedicated module for
560 Ground Glass Opacities and subtle nodules detection could further improve it. An iterative

561 optimisation of the training procedure, which would be possible when increasing the gold
562 standard dataset to be analysed, would also likely provide a better false positive rejection.

563 The main aim of the present study, which was to verify to what extent the M5L results
564 changed when progressively increasing the dataset size, was accomplished.

565 The performance is also independent of the input dataset, a feature that is rarely ad-
566 dressed in the literature: very similar results are obtained on *LIDC/IDRI*, *ANODE09* and
567 *ITALUNG-CT* scans.

568 The lungCAM overall performance is comparable and sometimes better than that of other
569 systems that were optimised on large and validated on small datasets, as opposite to our
570 strategy.

571 Excellent results are obtained when combining M5L to other systems, such as IsiCAD
572 and FlyerScan: A multi-thread CAD system based on the combination of several algorithms,
573 which could be made available thanks to WEB and cloud-based services, is indeed likely to
574 perform on standards that are compatible with those of an experienced radiologist and
575 would therefore provide a remarkable added value when used to support radiologists in
576 clinical practice and screening programs.

577 **ACKNOWLEDGMENTS**

578 The authors acknowledge the National Cancer Institute and the Foundation for the
579 National Institutes of Health and their role in the creation of the free publicly available
580 LIDC/IDRI Database used in this study. Thanks are due to Dr. F. Falaschi, Dr. C. Spinelli
581 and Dr. A. De Liperi (U.O. Radiodiagnostica 2, Azienda Ospedaliera Universitaria Pisana,
582 Pisa, Italy) for making available the dataset from the ITALUNG-CT trial. The invaluable
583 ANODE09 dataset with the associated radiologists annotations is available thanks to the
584 ANODE09 team, in particular prof. Bram van Ginneken who also compared the M5L results
585 with the ANODE09 annotations and provided the corresponding FROC curves.

586 **REFERENCES**

587 ¹D. Parkin, J. Ferlay, H. Shin, F. Bray, D. Forman, and C. Mathers, “Estimates of world-
588 wide burden of cancer in 2008,” *Int J Cancer* , 2893–2917 (2010).

589 ²American Cancer Society, “Cancer Facts and Figures.” (2009),
590 <http://www.cancer.org/Research/CancerFactsFigures>.

591 ³S. Diederich, M. Lentschig, T. Overbeck, D. Wormanns, and W. Heindel, “Detection of
592 pulmonary nodules at spiral CT: comparison of maximum intensity projection sliding slabs
593 and single-image reporting,” *European Radiology* **11**, 1345–1350 (2001).

594 ⁴C. A. van Iersel, H. J. de Koning, G. Draisma, W. P. T. M. Mali, E. T. Scholten, K. Nack-
595 aerts, M. Prokop, J. D. F. Habbema, M. Oudkerk, and R. J. van Klaveren, “Risk-based
596 selection from the general population in a screening trial: Selection criteria, recruitment
597 and power for the dutch-belgian randomised lung cancer multi-slice CT screening trial
598 (NELSON),” *International Journal of Cancer* **120**, 868–874 (2007).

599 ⁵A. L. Pegna, G. Picozzi, M. Mascalchi, F. M. Carozzi, L. Carrozzi, C. Comin, C. Spinelli,
600 F. Falaschi, M. Grazzini, F. Innocenti, C. Ronchi, and E. Paci, “Design, recruitment and
601 baseline results of the ITALUNG trial for lung cancer screening with low-dose CT,” *Lung
602 Cancer* **64**, 34–40 (2009).

603 ⁶National Lung Screening Trial Research Team, D. Aberle, C. Berg, W. Black, T. Church,
604 R. Fagerstrom, B. Galen, I. Gareen, C. Gatsonis, J. Goldin, J. Gohagan, B. Hillman,
605 C. Jaffe, B. Kramer, D. Lynch, P. Marcus, M. Schnall, D. Sullivan, D. Sullivan, and
606 C. Zylak, “The national lung screening trial: overview and study design,” *Radiology* **258**,
607 243–53 (2011).

608 ⁷National Lung Screening Trial Research Team, D. R. Aberle, A. M. Adams, C. D.
609 Berg, W. C. Black, R. M. Fagerstrom, I. F. Gareen, C. Gatsonis, P. M. Marcus,
610 and J. D. Sicks, “Reduced lung-cancer mortality with low-dose computed tomographic
611 screening,” *New England Journal of Medicine* **365**, 395–409 (2011), pMID: 21714641,
612 <http://www.nejm.org/doi/pdf/10.1056/NEJMoa1102873>.

613 ⁸H. Roberts, D. Patsios, D. Kucharczyk, N. Paul, and T. Roberts, “The utility of computer-
614 aided detection (CAD) for lung cancer screening using low-dose CT, proceedings of the
615 19th international congress and exhibition, Berlin, June 22 - 25,” *International Congress
616 Series* , 1137–1142 (2005).

617 ⁹M. Das, G. Mühlenbruch, A. Mahnken, T. Flohr, L. Gündel, S. Stanzel, T. Kraus, r. R.
618 Günthe, and J. Wildberger, “Small pulmonary nodules: effect of two computer-aided
619 detection systems on radiologist performance.” *Radiology* **241**, 564–571 (2006).

- 620 ¹⁰B. Brochu, C. Beigelman-Aubry, J. Goldmard, P. Raffy, P. Grenier, and O. Lucidarme,
621 “Computer-aided detection of lung nodules on thin collimation MDCT: impact on radiol-
622 ogists performance.” *Journal de Radiologie* **88**, 573–578 (2007).
- 623 ¹¹S. Matsumoto, Y. Ohno, H. Yamagata, D. Takenaka, and K. Sugimura, “Computer-
624 aided detection of lung nodules on multidetector row computed tomography using three-
625 dimensional analysis of nodule candidates and their surroundings.” *Radiation Medicine*
626 **26**, 562–569 (2008).
- 627 ¹²M. S. Brown, J. G. Goldin, S. Rogers, H. J. Kim, R. D. Suh, M. F. McNitt-Gray, S. K.
628 Shah, D. Truong, K. Brown, J. W. Sayre, D. W. Gjertson, P. Batra, and D. R. Aberle,
629 “Computer-aided lung nodule detection in CT: Results of large-scale observer test1,” *Aca-
630 demic Radiology* **12**, 681–686 (2005).
- 631 ¹³B. Sahiner, H.-P. Chan, L. M. Hadjiiski, P. N. Cascade, E. A. Kazerooni, A. R. Chughtai,
632 C. Poopat, T. Song, L. Frank, J. Stojanovska, and A. Attili, “Effect of CAD on radiologists’
633 detection of lung nodules on thoracic CT scans: Analysis of an observer performance study
634 by nodule size,” *Academic Radiology* **16**, 1518–1530 (2009).
- 635 ¹⁴A. Retico, P. Delogu, M. E. Fantacci, I. Gori, and A. P. Martinez, “Lung nodule detection
636 in low-dose and thin-slice computed tomography,” *Computers in biology and medicine* **38**,
637 525–534 (2008).
- 638 ¹⁵A. Retico, M. E. Fantacci, I. Gori, P. Kasae, B. Golosio, A. Piccioli, P. Cerello, G. D.
639 Nunzio, and S. Tangaro, “Pleural nodule identification in low-dose and thin-slice lung
640 computed tomography,” *Computers in biology and medicine* **39**, 1137–1144 (2009).
- 641 ¹⁶S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves,
642 B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon,
643 E. J. R. van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M.
644 Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P.-Y. Qing, R. Y. Roberts, A. R.
645 Smith, A. Starkey, P. Batra, and P. Caligiuri, “The lung image database consortium
646 (LIDC) and image database resource initiative (IDRI): A completed reference database of
647 lung nodules on CT scans,” *Medical Physics* **38**, 915–931 (2011).
- 648 ¹⁷B. van Ginneken, S. G. Armato III, B. de Hoop, S. van Amelsvoort-van de Vorst, T. Duin-
649 dam, M. Niemeijer, K. Murphy, A. Schilham, A. Retico, M. E. Fantacci, N. Camarlinghi,
650 F. Bagagli, I. Gori, T. Hara, H. Fujita, G. Gargano, R. Bellotti, S. Tangaro, L. Bolanos,
651 F. D. Carlo, P. Cerello, S. C. Cheran, E. L. Torres, and M. Prokop, “Comparing and

652 combining algorithms for computer-aided detection of pulmonary nodules in computed
653 tomography scans: The ANODE09 study,” *Medical Image Analysis* **14**, 707–722 (2010).

654 ¹⁸N. Camarlinghi, I. Gori, A. Retico, R. Bellotti, P. Bosco, P. Cerello, G. Gargano, E. L.
655 Torres, R. Megna, M. Peccarisi, and M. E. Fantacci, “Combination of computer-aided
656 detection algorithms for automatic lung nodule identification,” *International Journal of*
657 *Computer Assisted Radiology and Surgery* **7**, 455–464 (2012).

658 ¹⁹R. Bellotti, P. Cerello, S. Tangaro, V. Bevilacqua, M. Castellano, G. Mastronardi, F. D.
659 Carlo, S. Bagnasco, U. Bottigli, R. Cataldo, E. Catanzariti, S. C. Cheran, P. Delogu, I. D.
660 Mitri, G. D. Nunzio, M. E. Fantacci, F. Fauci, G. Gargano, B. Golosio, P. L. Indovina,
661 A. Lauria, E. L. Torres, R. Magro, G. L. Masala, R. Massafra, P. Oliva, and A. P. Mar-
662 tinez, “Distributed medical images analysis on a Grid infrastructure,” *Future Generation*
663 *Computer Systems* **23**, 475–484 (2007).

664 ²⁰R. Bellotti, F. D. Carlo, G. Gargano, S. Tangaro, D. Cascio, E. Catanzariti, P. Cerello,
665 S. C. Cheran, P. Delogu, I. D. Mitri, C. Fulcheri, D. Grosso, A. Retico, S. Squarcia,
666 E. Tommasi, and B. Golosio, “A CAD system for nodule detection in low-dose lung CTs
667 based on region growing and a new active contour model,” *Medical Physics* **34**, 4901–4910
668 (2007).

669 ²¹A. Savitzky and M. J. E. Golay, “Smoothing and differentiation of data by simplified least
670 squares procedures.” *Analytical Chemistry* **36**, 1627–1639 (1964).

671 ²²S. Rajagopalan and R. A. Robb, “Image smoothing with Savtizky-Golay filters. Proc. SPIE
672 5029, *Medical Imaging 2003: Visualization, image-guided procedures*,” , 773–781 (2003).

673 ²³G. D. Nunzio, E. Tommasi, A. Agrusti, R. Cataldo, I. D. Mitri, M. Favetta, S. Maglio,
674 A. Massafra, M. Quarta, M. Torsello, I. Zecca, R. Bellotti, S. Tangaro, P. Calvini, N. Ca-
675 marlinghi, F. Falaschi, P. Cerello, and P. Oliva, “Automatic lung segmentation in CT
676 images with accurate handling of the hilar region,” *Journal of Digital Imaging* **24**, 11–27
677 (2011).

678 ²⁴P. Cerello, S. C. Cheran, S. Bagnasco, R. Bellotti, L. Bolanos, E. Catanzariti, G. D. Nunzio,
679 M. E. Fantacci, E. Fiorina, G. Gargano, G. Gemme, E. L. Torres, G. L. Masala, C. Peroni,
680 and M. Santoro, “3-D object segmentation using ant colonies,” *Pattern Recognition* **43**,
681 1476–1490 (2010).

682 ²⁵T. Messay, R. C. Hardie, and S. K. Rogers, “A new computationally efficient CAD system
683 for pulmonary nodule detection in CT imagery,” *Medical Image Analysis* **14**, 390–406

684 (2010).

685 ²⁶S. L. A. Lee, A. Z. Kouzani, and E. J. Hu, “Automated detection of lung nodules in
686 computed tomography images: a review,” *Machine Vision and Applications* **23**, 151–163
687 (2012).

688 ²⁷H. S. Pheng, S. M. Shamsuddin, and S. Kenji., “Application of intelligent computational
689 models on computed tomography lung images,” *International Journal of Advances in Soft
690 Computing and its Applications* **3**, 1–15 (2011).

691 ²⁸M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, “A novel computer-aided lung
692 nodule detection system for ct images,” *Medical Physics* **38**, 5630–5645 (2011).

693 ²⁹B. Golosio, G. L. Masala, A. Piccioli, P. Oliva, M. Carpinelli, R. Cataldo, P. Cerello,
694 F. De Carlo, F. Falaschi, M. E. Fantacci, G. Gargano, P. Kasae, and M. Torsello, “A
695 novel multithreshold method for nodule detection in lung ct,” *Medical Physics* **36**, 3607–
696 3618 (2009).

697 ³⁰W. Guo and Q. Li, “High performance lung nodule detection schemes in ct using local and
698 global information,” *Medical Physics* **39**, 5157–5168 (2012).

699 ³¹M. Brown, P. Lo, J. Goldin, E. Barnoy, G. Kim, M. McNitt-Gray, and D. Aberle, “Toward
700 clinically usable cad for lung cancer screening with computed tomography,” *European
701 Radiology* , 1–10 (2014).

702 ³²K. Murphy, B. van Ginneken, A. M. R. Schilham, B. J. de Hoop, H. A. Gietema, and
703 M. Prokop, “A large-scale evaluation of automatic pulmonary nodule detection in chest CT
704 using local image features and k-nearest-neighbour classification,” *Medical Image Analysis*
705 **13**, 757–770 (2009), includes Special Section on the 12th International Conference on
706 Medical Imaging and Computer Assisted Intervention.

707 ³³M. Godoy, P. Cooperberg, Z. Maizlin, R. Yuan, A. McWilliams, S. Lam, and J. Mayo,
708 “Detection sensitivity of a commercial lung nodule cad system in a series of pathologically
709 proven lung cancers.” *Journal of Thoracic Imaging* **23**, 1–6 (2008).

710 ³⁴R2 Technology, Inc, “Understanding the imagechecker ct lung system. pn 13229 rev a,”
711 (Hologic Inc; 35 Crosby Drive, Bedford, MA 01730-01401, 2005).

DICOM
IMAGE

Preprocessor

3-D Lung
Segmentation

Image Filtering
(Savitzky-Golay)

Lung
mask

3-D Candidates
Detection and
Segmentation

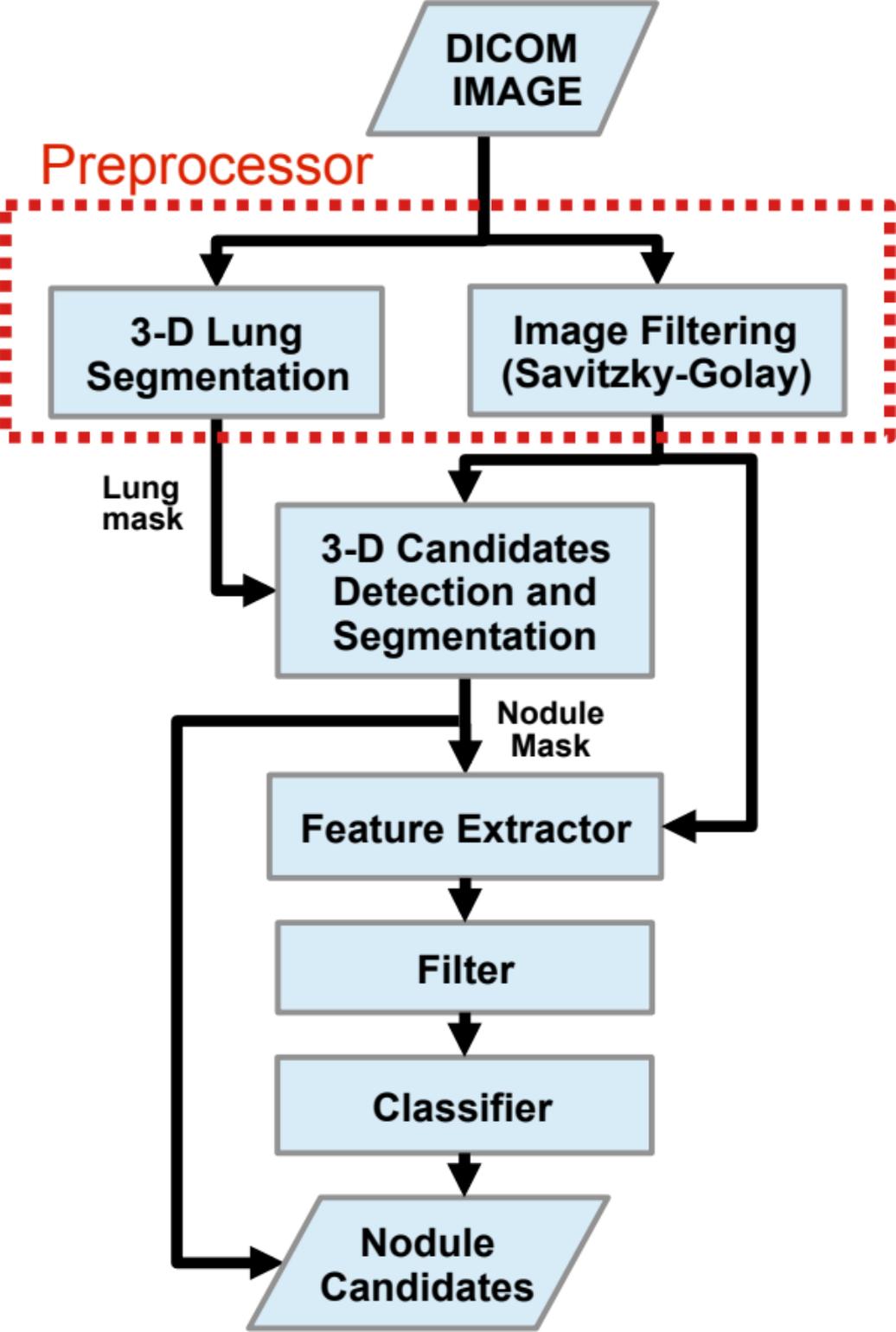
Nodule
Mask

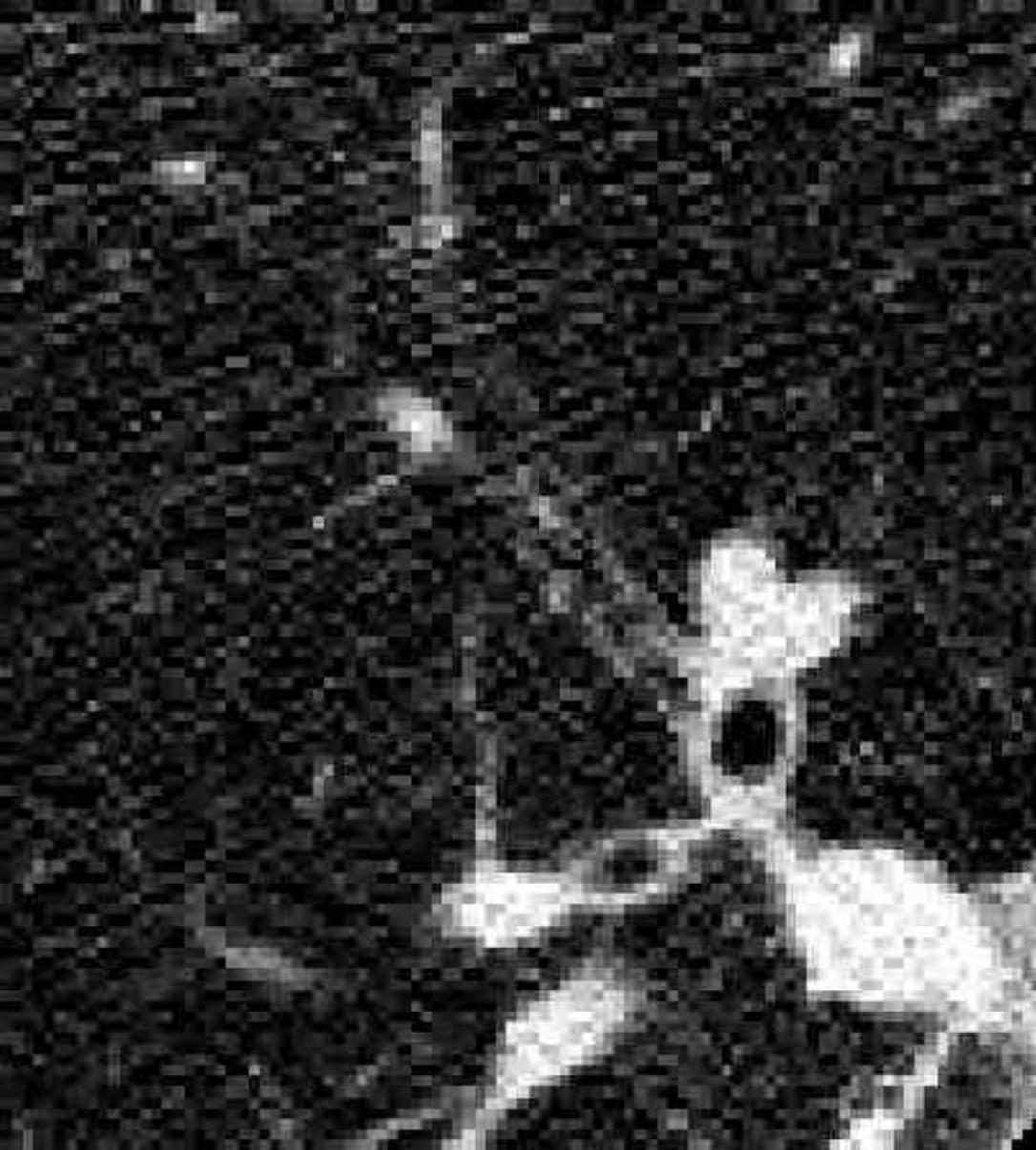
Feature Extractor

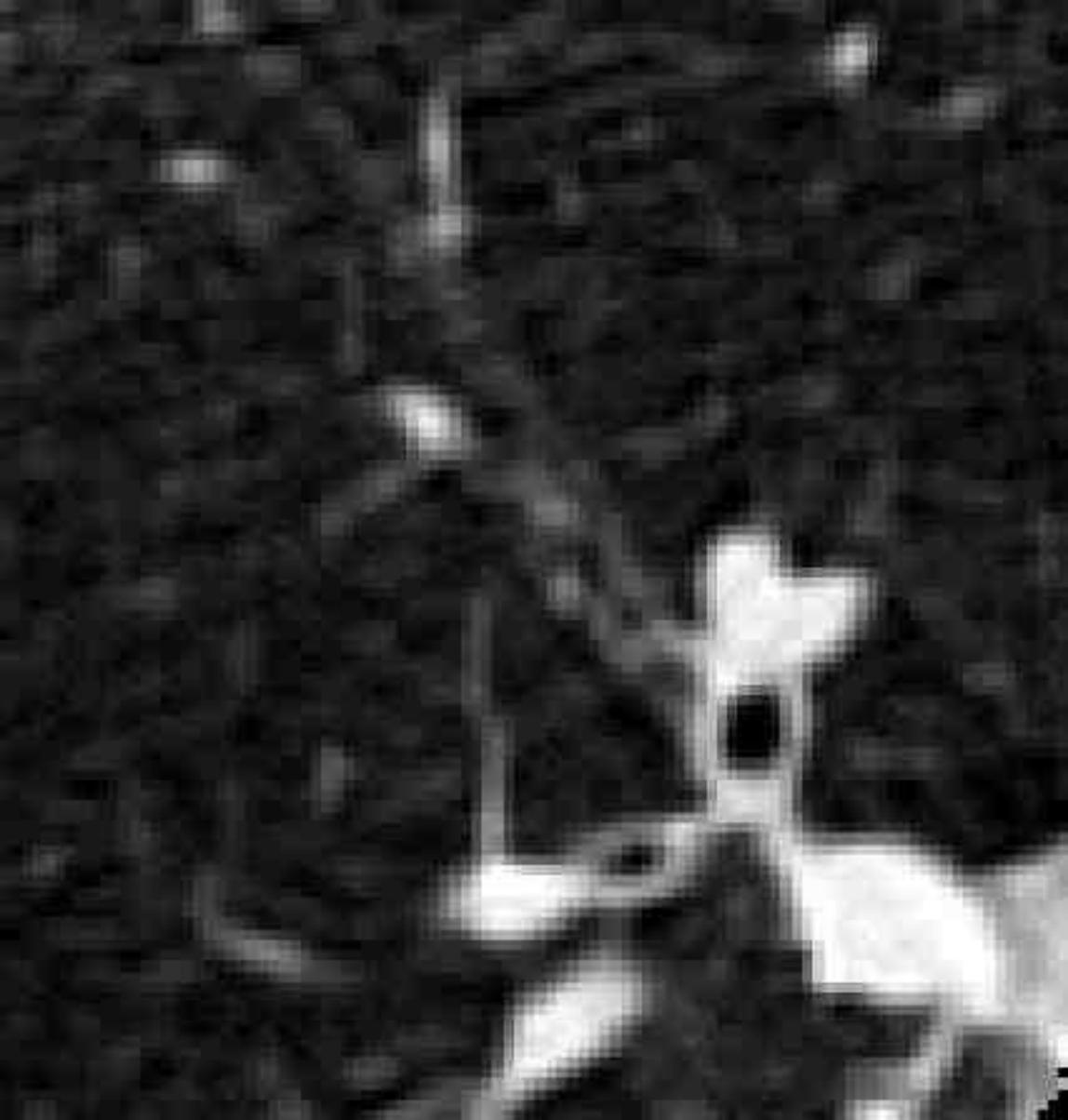
Filter

Classifier

Nodule
Candidates



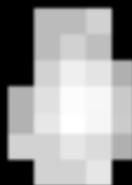


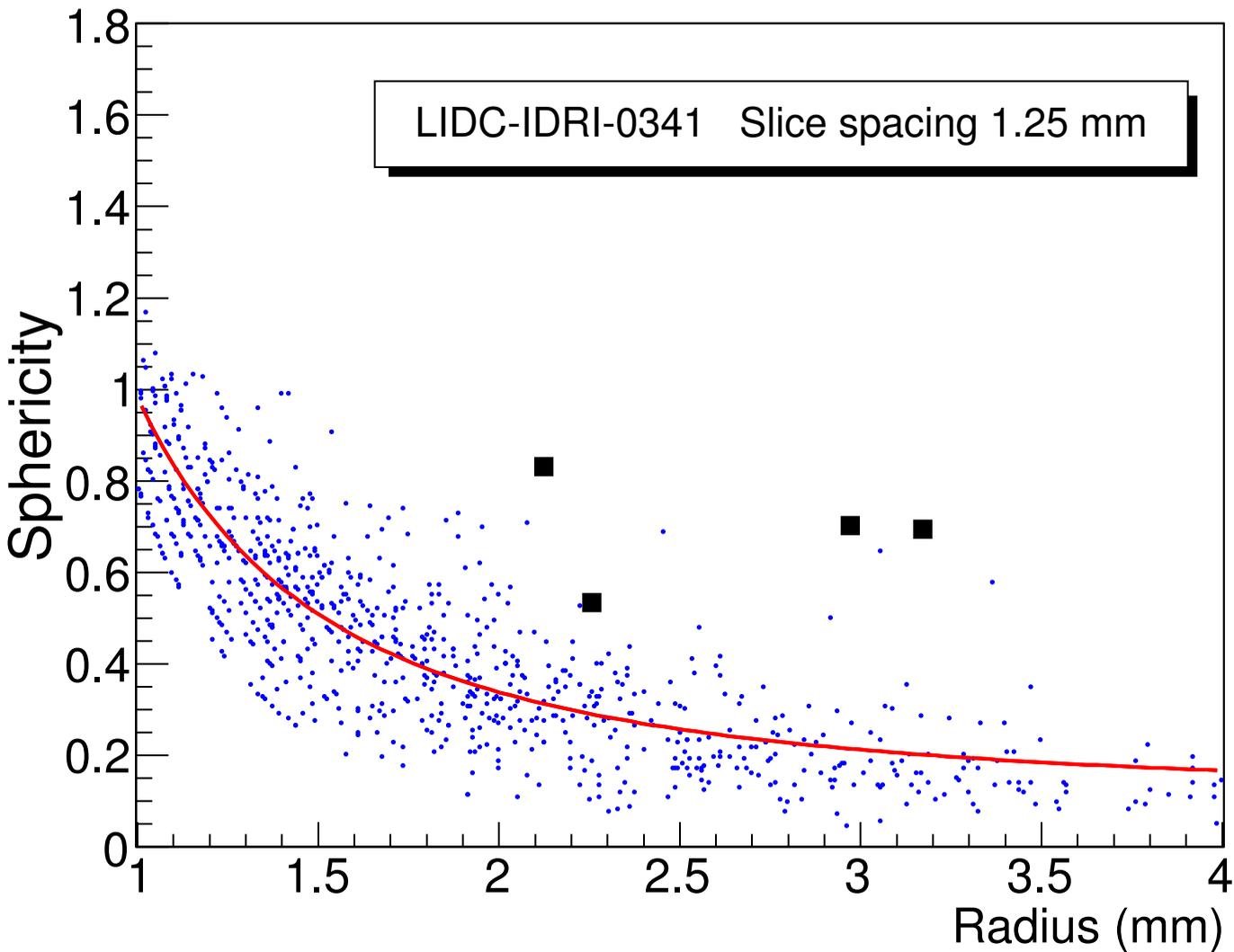


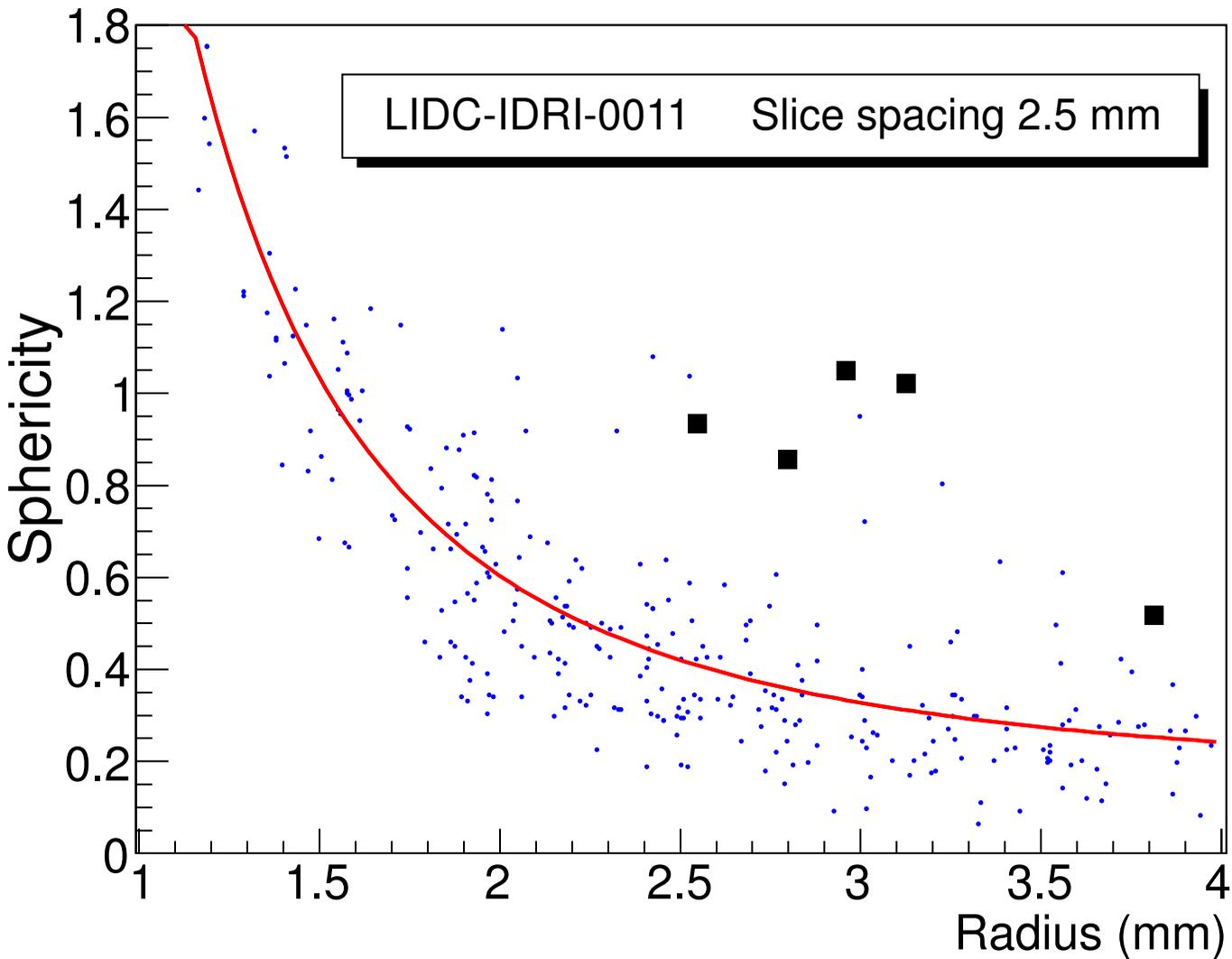


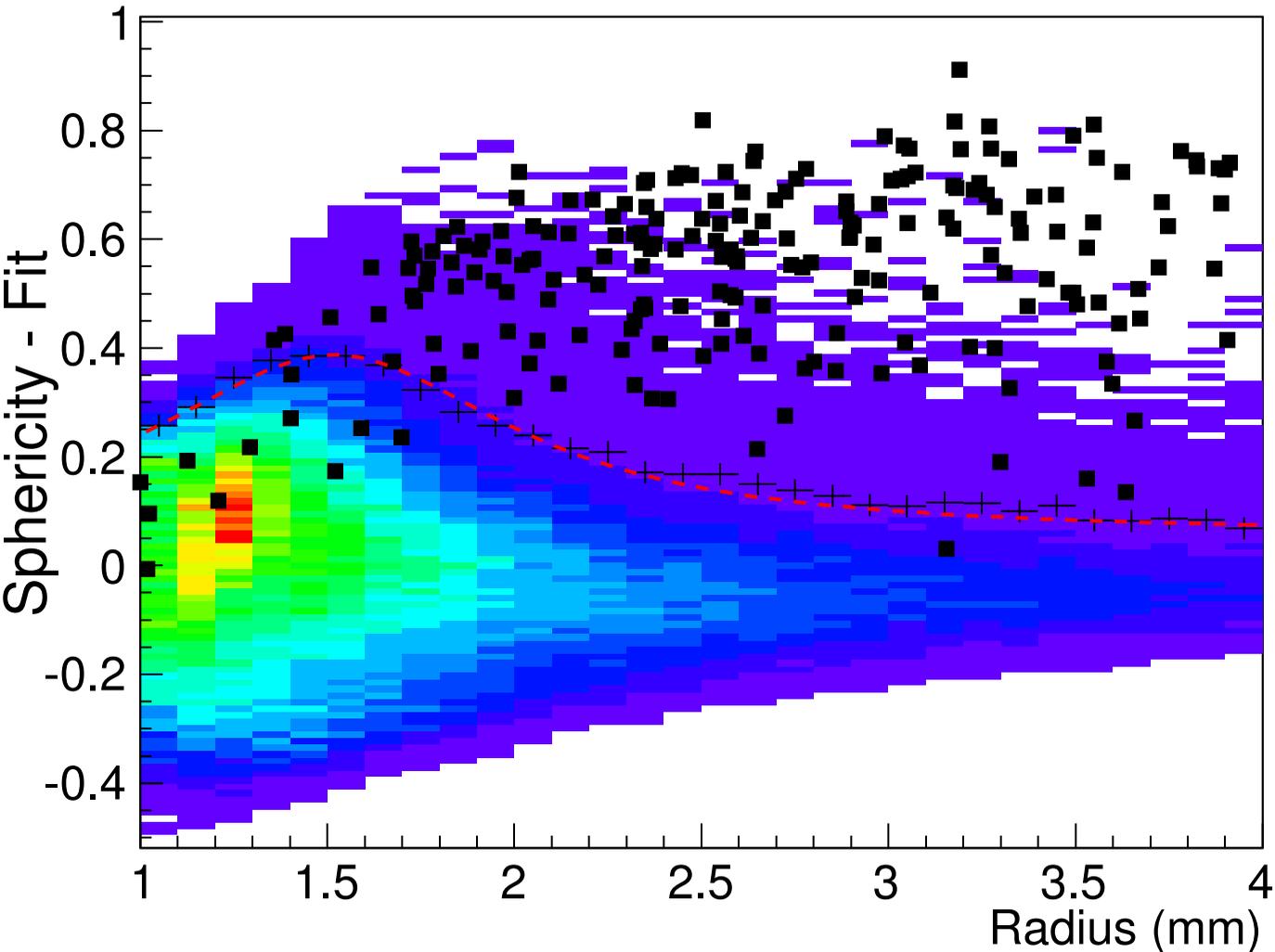




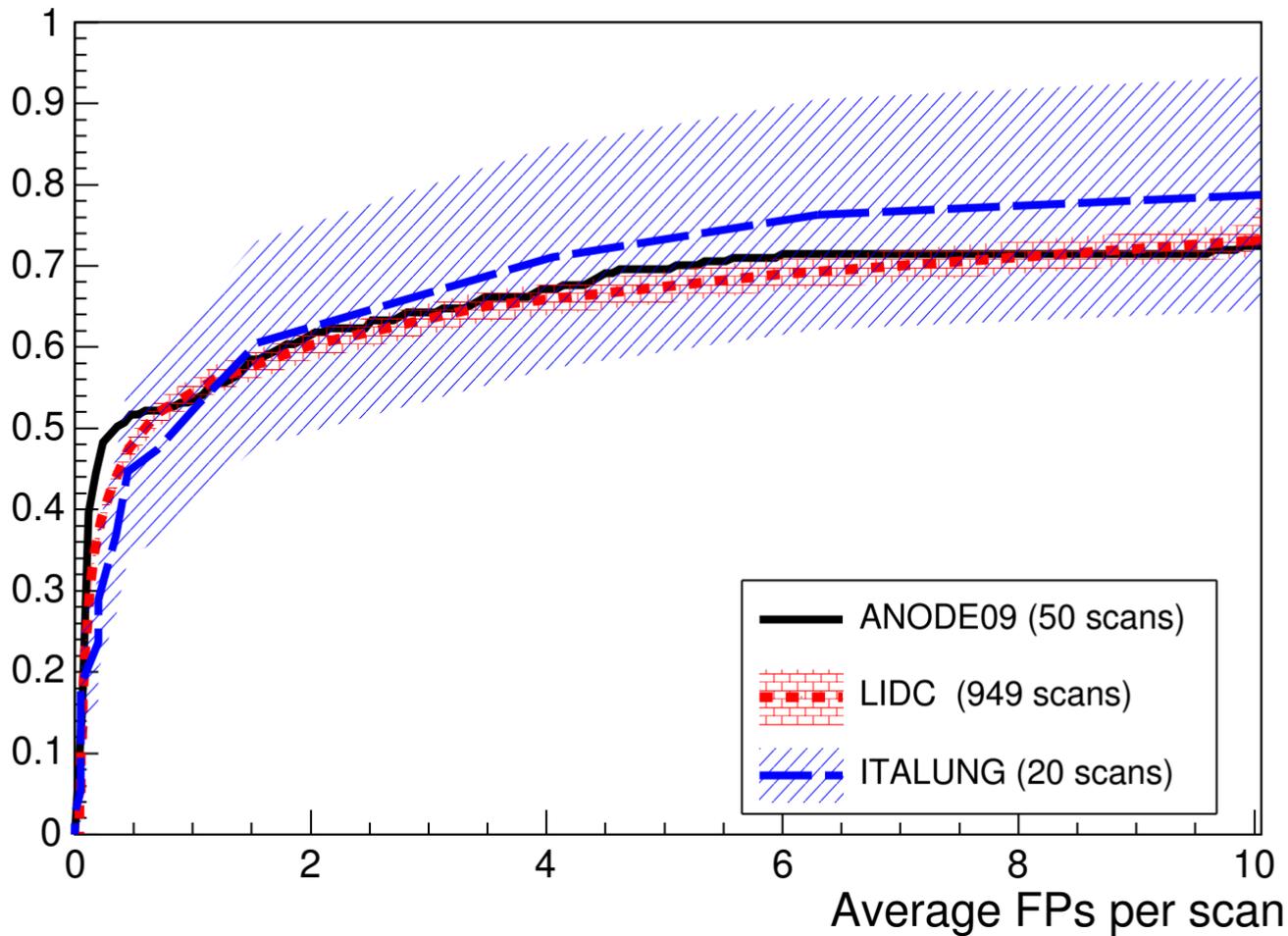






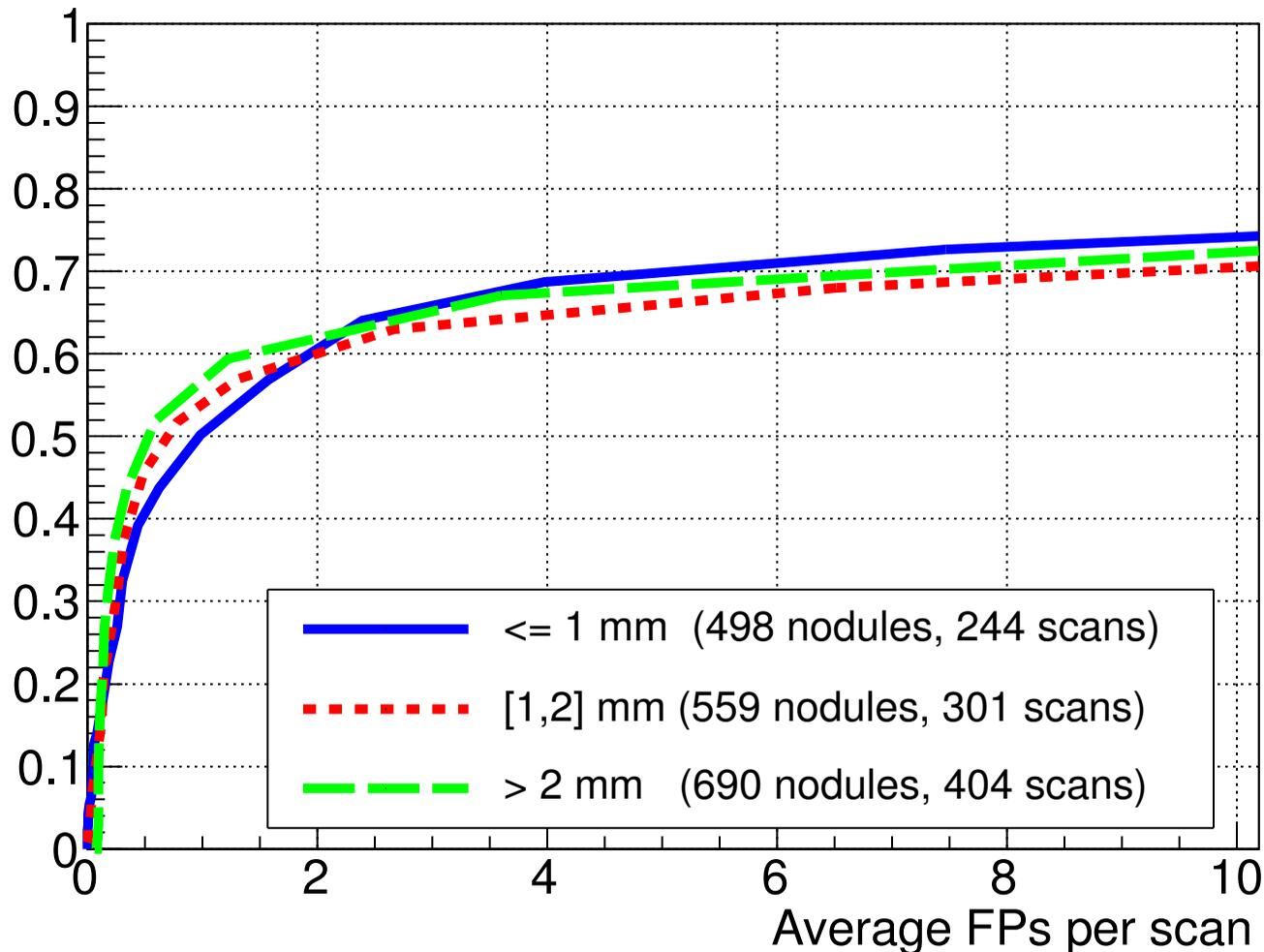


Sensitivity

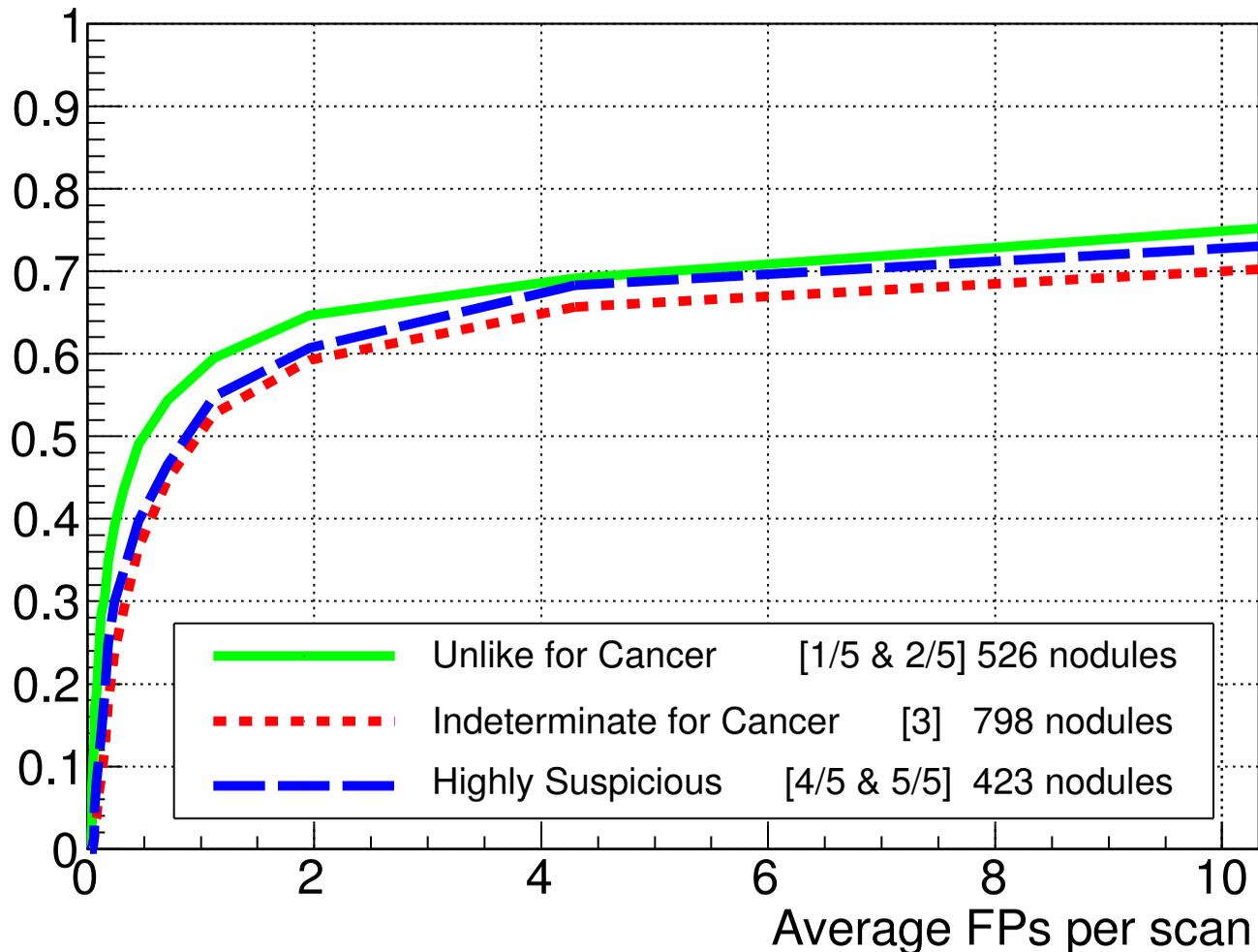


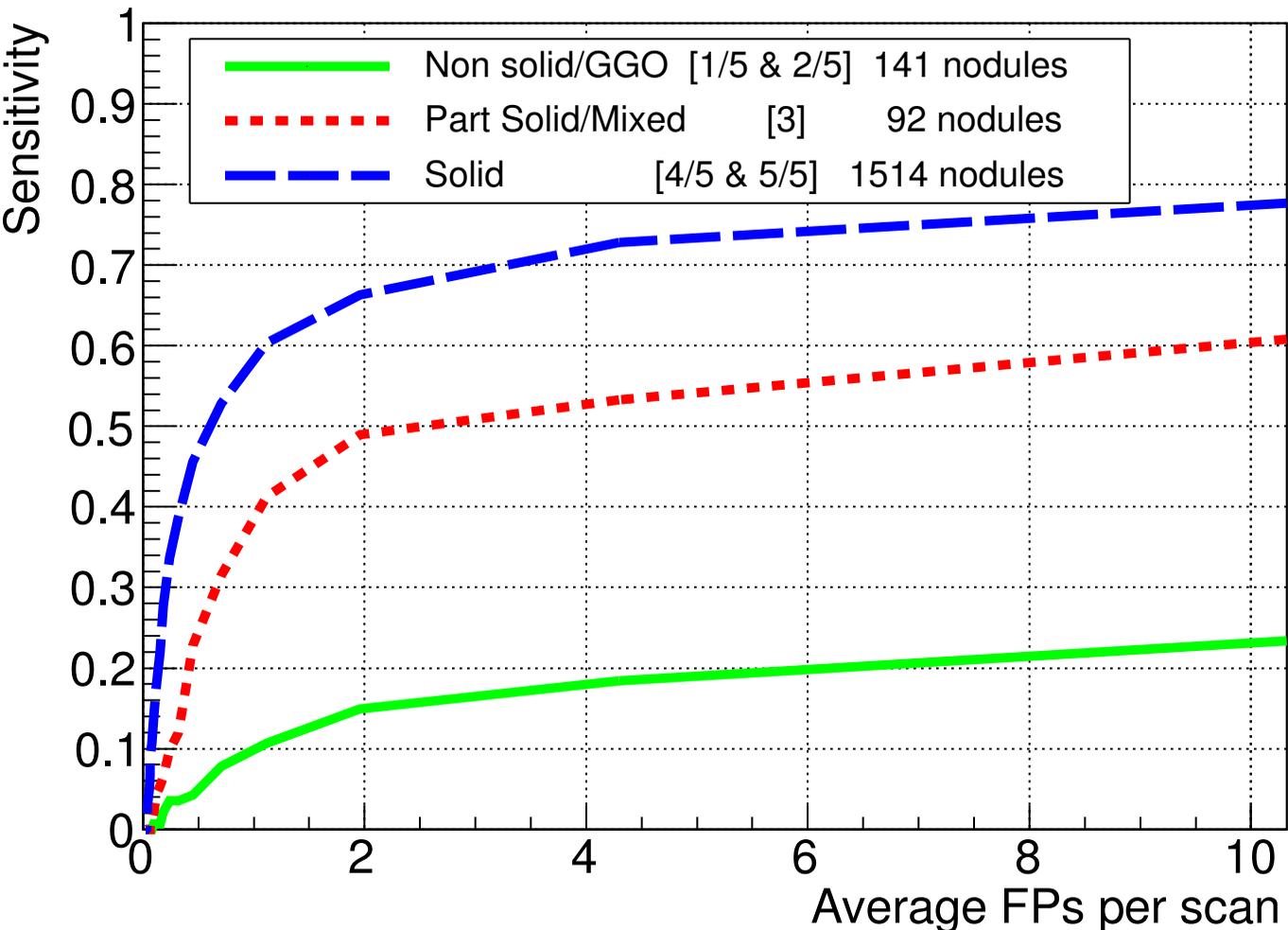
- ANODE09 (50 scans)
- - LIDC (949 scans)
- - ITALUNG (20 scans)

Sensitivity



Sensitivity





Sensitivity

