

A Web- and Cloud- based Service for the Clinical Use of a CAD (Computer Aided Detection) System

Automated Detection of Lung Nodules in Thoracic CTs (Computed Tomographies)

M. E. Fantacci^{1,2}, A. Traverso^{3,4}, S. Bagnasco⁴, C. Bracco⁵, D. Campanella⁶, G. Chiara⁶,
E. Lopez Torres^{4,7}, A. Manca⁶, D. Regge⁶, M. Saletta⁴, M. Stasi⁵, S. Vallero⁴, L. Vassallo⁶ and P. Cerello⁴

¹*Department of Physics, University of Pisa, Largo Pontecorvo 3, Pisa, Italy*

²*INFN, Sezione di Pisa, Pisa, Italy*

³*Department of Applied Science and Technology, Polytechnic University of Turin, Turin, Italy*

⁴*INFN, Sezione di Torino, Turin, Italy*

⁵*Medical Physics Unit, Candiolo Cancer Institute - FPO, IRCCS, Candiolo, Italy*

⁶*Radiology Unit, Candiolo Cancer Institute - FPO, IRCCS, Candiolo, Italy*

⁷*CEADEN, Havana, Cuba*

Keywords: Web Service, Cloud Computing, Computer Aided Detection, Lung Nodules.

Abstract: M5L, a Web-based Computer-Aided Detection (CAD) system to automatically detect lung nodules in thoracic Computed Tomographies, is based on a multi-thread analysis by independent subsystems and the combination of their results. The validation on 1043 scans of 3 independent data-sets showed consistency across data-sets, with a sensitivity of about 80% in the 4-8 range of False Positives per scan, despite varying acquisition and reconstruction parameters and annotation criteria. To make M5L CAD available to users without hardware or software new installations and configuration, a Software as a Service (SaaS) approach was adopted. A web front-end handles the work (image upload, results notification and direct on-line annotation by radiologists) and the communication with the OpenNebula-based cloud infrastructure, that allocates virtual computing and storage resources. The exams uploaded through the web interface are anonymised and analysis is performed in an isolated and independent cloud environment. The average processing time for case is about 20 minutes and up to 14 cases can be processed in parallel. Preliminary results on the on-going clinical validation shows that the M5L CAD adds 20% more nodules originally overlooked by radiologists, allowing a remarkable increase of the overall detection sensitivity.

1 INTRODUCTION

Lung Cancer is one of the main health issues in developed countries, accounting for about 20% and 28% of cancer-related deaths in Europe (Malvezzi, 2015) and the United States of America (Society, 2015), respectively, with a five-year survival rate of only 10-17% (Society, 2015). Computed Tomography (CT) has been shown to be the most sensitive imaging modality for the detection of small pulmonary nodules, which constitutes the first radiological sign of this pathology. Low dose high resolution CT-based screening trials are regarded as a promising technique for detecting early-stage lung

cancers (National Lung Screening Trial, 2011; Moyer, 2014). However, the identification of early stage pathological Regions of Interest (ROIs) in low-dose high-resolution CT scans is a difficult and time consuming task for radiologists, because of the large number (300/500) of noisy 2D slices to be analysed. In order to support radiologists, researchers have developed CAD algorithms to be applied to CT scans. Several studies (Das, 2006; Brochu, 2007; Matsumoto, 2008) reported an improvement in the sensitivity of radiologists when assisted by CAD systems, which also act as detection rates equalisers between observers of different level of experience (Brown, 2005). The most common approach to make

CAD algorithms available in the clinical routine is the deployment of standalone workstations, usually equipped with a vendor-dependent Graphic User Interface (GUI), which presents several drawbacks: the high fixed cost of software licenses, a dedicated hardware, their rapid obsolescence. Furthermore, the computational needs of CAD algorithms, depending on their complexity, often require powerful and expensive hardware. The diffusion of Cloud Computing solutions, accessible via secure Web protocols, solves almost all the previous issues. In addition, the SaaS approach provides the possibility of combining several CADs, with demonstrated benefits to the overall performance (van Ginneken, 2010). This paper presents the M5L on-demand CAD system: Section 2 summaries the main features of the algorithms and the results of their validation, describes the functional blocks of the M5L on-demand service (the web front-end and the cloud back-end) and discusses the methodology that was adopted to perform stress test on the developed system and the on-going clinical validation; Section 3 discusses the results of the stress test and the preliminary results of the clinical validation; Section 4 draws some conclusions and analyses possible further developments.

2 MATERIALS AND METHODS

2.1 The M5L CAD System

M5L is the combination of two independent CAD sub-systems: the Channeler Ant Model (lungCAM) and the Voxel-Based Neural Approach (VBNA). These two algorithms have a common starting point, which is the parenchymal volume, obtained with a 3D region growing segmentation algorithm, that also excludes the trachea and separates the two lungs (De Nunzio, 2011).

2.1.1 The LungCAM CAD

This algorithm is based on the reproduction of the life-cycle of colonies of virtual ants (Cerello, 2008), released from an anthill with the capability to move along the 3D space determined by the lung volume in the CT. The motion is accompanied by the release of pheromone traces along the ant path. The CT voxel intensity is interpreted as the amount of food available to the ants and is progressively reduced by the ant feeding. The evolution of the colony is determined by a set of rules, which define how ants move, the released amount of pheromone and the

cycles of reproduction and death. The algorithm ends when all the ants in the colony have died. The output of this stage is a pheromone map, a collection of segmented objects, which are classified by means of a feed-forward artificial neural network with 13 input features, which take into account both geometrical (as for example radius, sphericity, skewness, kurtosis) and intensity (as for example average, standard deviation and entropy inside and outside the mask, maximum) characteristics of the segmented objects, as described in detail in (Lopez Torres, 2015). The lungCAM algorithm detects both pleural and internal nodules.

2.1.2 The VBNA CAD

This algorithm makes use of two different procedures to detect nodules inside the lung parenchyma (CADI) (Li, 2003; Retico, 2008) and nodules attached to the pleura (CADJP) (Retico, 2009). The former method models nodules as spheres with a Gaussian profile, where the centre of the sphere is chosen to be corresponding to a local maximum in the voxel intensity. The CADJP method builds normals to the pleura surface and each voxel gets a score depending on the number of normals crossing it. Before combining the CADI and CADJP results, nodule candidates are classified by a Support Vector Machine (SVM) using more than 100 voxel-based input features for each voxel (the intensity values of its 3D neighbors, in particular the $5 \times 5 \times 5$ intensity values, and the eigenvalues of the gradient matrix and of the Hessian matrix). At the end of the procedure, each ROI is assigned a degree of suspicion averaging the score of all the voxels belonging to it (Camarlinghi, 2012).

2.1.3 M5L Results

The M5L results, obtained combining lungCAM and VBNA have been evaluated in terms of FROC (Free Response Receiver Operating) curves, as described in (Lopez Torres, 2015). The measured M5L sensitivity, evaluated on several input data-sets, including both the ANODE09 dataset (van Ginneken, 2010) and the full LIDC/IDRI database (Armato, 2011), reaches 80% at 4-8 FP/scan, which, given the size (more than 1000 scans) and heterogeneity of the data-sets, is satisfactory.

2.2 The M5L on Demand System

In order to make M5L easily available to radiologists without any additional hardware or software requirement, a Web-based interface has

been designed and implemented. Furthermore, in order to efficiently exploit computational resources, a Cloud-computing facility was setup. The M5L on-demand system is then composed by the following modules: the Web front-end, designed to provide radiologists with the functionality to exchange imaging studies and compare diagnoses on the same studies, handles the CT submission, the on-line insertion of the medical review and the access to CAD results; the Cloud Computing back-end, conceived to optimize the access to the available computing resources (i.e. virtual machines) according to the user requests, handles the algorithms execution. A similar concept was also recently proposed as possible future development (Goo, 2015).

2.2.1 The Web Front-end

The M5L service is available as a web application accessible with any browser from desktop and mobile devices. With proper credentials, users can access the service and upload thoracic CT studies in DICOM format to the remote repository; reviewers (not necessarily belonging to the same institution) can insert their findings. The notification system is based on emails, that alert reviewers when a new case has to be annotated and when the M5L results are ready. The entire front-end web interface has been developed using the DRUPAL free and open source tool, a content management platform powering millions of websites and applications (Coomb, 2009). The DRUPAL's core is modular, defining a system of hooks and callbacks which are accessed internally via Application Programming Interface (API). This design allows the integration of third-party contributed modules and themes to extend and/or override DRUPALs default behaviours without changing its core code. DRUPAL easily allows the development of custom modules in PHP code. In terms of security and data access, DRUPAL offers the possibility to the site administrator to define group of users, with different credentials and this allows, according to user requests, to protect content access. Two different modules and the associated user profiles have been developed, which are briefly described in the following parts. The submission module is conceived to be used by a technician operating with a submitter role, who uploads one or more CT studies to be analysed and selects one or more radiologists (not necessarily belonging to the same institutions) who will review the studies. The module layout is shown in Figure 1. Users can

submit studies sequentially or asynchronously: an email will inform them when the M5L analysis and the review by radiologists are completed. Before being submitted for computation, all cases are subject to further validation checks (e.g. image quality controls) that include a re-anonymisation. The review module is conceived to be used by a radiologist. After logging in with reviewer role, a radiologist can insert the medical annotation of studies assigned for review during the submission process. The M5L results are available in different formats, such as DICOM Structured Report, HTML, XML and PDF. A custom plug-in for the Osirix medical imaging viewer (Rosset, 2004) has also been developed to view CAD marks directly on the CT scan. The radiologist can operate as *first* or *second reader*. When operating as *first reader*, the M5L results are unlocked only after the annotation is completed and validated. The M5L marks can then be reviewed and assessed, with the options to include them in the annotation, as shown in Figure 2, reject them as false positive findings or label them as *non-relevant* nodules.

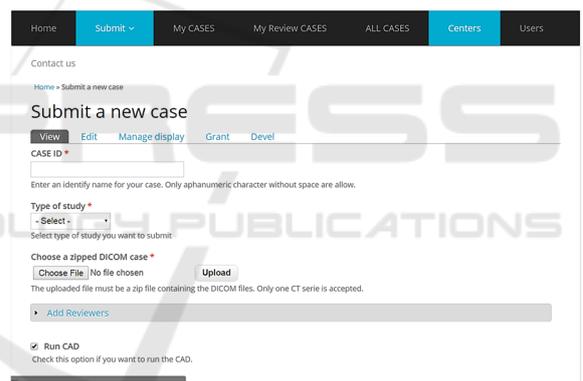


Figure 1: Module of the Web interface for the CT submission to the M5L CAD. Setting the various fields it is possible to select the exam and the radiologist that will be asked to review it.

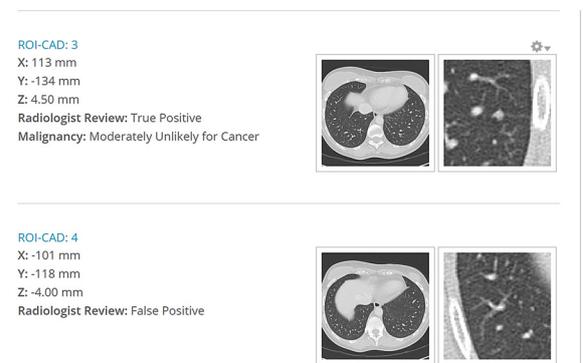


Figure 2: Examples of candidate nodules in the M5L CAD output and the associated evaluation by the radiologist.

In order to speed-up the overall double reading time, an automated matching algorithm compares the CAD marks with annotated ROIs. The radiologist can then concentrate only on overlooked findings. When a radiologist operates as *second-reader*, the M5L marks are immediately available.

2.2.2 The Cloud Back-end

One of the issues when running CAD algorithms in parallel and combining their results is that the required memory and computing power is large and not precisely predictable. Furthermore, if the tool is to be used in potentially massive screening campaigns, which can be concentrated in a given period of the year, some flexibility is required in the allocation of computing resources. In the past, prototype systems for handling the analysis of medical images in a distributed environment based on the use of a GRID infrastructure, used in high-energy physics, were deployed (Bellotti, 2007). However, the main issues with Grid Computing are its rigidity, the complexity of the distributed structure and the manpower required to manage the system. This solution does not fit well in the case in the majority of Medical Physics projects, where a custom environment must be configured according to the requirements set by the users. Furthermore, the management of such infrastructure requires additional man-power, making the solution not cost-effective for the hospital. Cloud Computing on the other hand is a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable and dynamically accessible computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell, 2011). Resources used by the different cloud applications are handled in completely separated environments (sandboxes), so it is impossible for a user to access the data of another or create damage elsewhere. Even Cloud administrators cannot access users resources, once created and properly configured. These features improve the overall security, privacy and data preservation. A Cloud Computing infrastructure can provide functionality to the users at several levels: basic infrastructure of network, storage and physical or virtual machines (Infrastructure As A Service or IaaS); computing platform with programs and services (Platform as a Service or PaaS); access to software solutions (Software As A Service or SaaS). One of the key Cloud Computing features, as highlighted in (Mell,

2011), is the elasticity of resources allocation, i.e. the capability to scale (up or down) the resources according to computing power needs. This point is even more important in case of a research Community Cloud, where the users have free access to resources that are shared between many groups and projects. The M5L CAD on-demand service is hosted by the INFN Torino Computing Centre, which is a Tier-2 of Large Hadron Collider Computing Grid (Turner, 2006). In order to reduce the manpower to manage the Tier-2, to support an increasing number of applications and to offer computational power to small research projects with specific needs, a Cloud infrastructure has been deployed (Bagnasco, 2014). The facility is managed by OpenNebula, a free and open-source Cloud Management Platform which allows hardware and virtual infrastructure control and monitoring, adding the possibility of virtual machine life-cycle management (Milojicic, 2011). OpenNebula orchestrates storage, networking, virtualization, monitoring, and security technologies to deploy computing services as virtual machines on distributed resources. The prototype M5L version makes use of one physical host as web-server and several Virtual Machines (VMs) as computational power. The VMs are deployed in a specific sandbox (IaaS) within an internal private network and a virtual router with a public address for the external access and the communication with the web server (that has another public IP). Presently, M5L is allowed to deploy up to 18 VMs, with a total of 48 cores and 100 GB persistent iSCSI (Internet Small Computer System Interface) storage disk, exported between all VMs via NFS (Network File System) and used for temporary storage of input CTs and the M5L output results. An elastic cluster based on CernVM Online (Buncic, 2010), a service that can create clusters with a head node and many workers based on CernVM OS (CERNVM, 2015), was configured in order to achieve the capability to scale resources up or down. VMs can be contextualised using CernVM Online to define the use of resources, user settings and automatically install and configure HTCondor and ElastiQ. HTCondor is a specialised work-load management system (batch system) for compute-intensive jobs (Tannenbaum, 2001): it accepts the submission of serial or parallel jobs (which are placed into a queue and run according to the selected policy), monitors their progress, and finally informs the user upon completion. ElastiQ (Berzano, 2014) is a lightweight Python daemon that allows a cluster of Virtual Machines running a batch system to scale up and down automatically. When a

CT is uploaded to the web service, it is copied to the head node of the cluster and the M5L CAD execution is controlled by HTCCondor. If some CTs to be examined are waiting in the HTCCondor queue more than a predefined period of time, new workers are deployed up to the limit of resources. Conversely if a worker is idle for more than a predefined period of time, it is de-allocated. In this way when no exam is uploaded only a small part of the available resources is locked, freeing the remaining for other Cloud users. Users are automatically alerted when the computations are completed, by means of a notification system based on e-mails. Since the M5L CAD is the combination of two independent CAD subsystems, for each exam two different HTCCondor jobs (one for the VBNA and one for the CAM CAD) run in parallel. When the analysis is completed for both CADs, the combination is made. HTCCondor is configured to run a maximum of two jobs at the same time for each worker node, otherwise it would assign jobs up to the number of cores of the virtual machine. This configuration keeps some cores free to use multi-threading (i.e. to spread a process over several cores) speeding up the analysis.

2.3 Stress Test Configuration

A stress test was set up to test the performance and the stability of the system under workloads comparable with daily clinical practice. Exams are submitted to M5L in two phases, using a custom python script: first simulating a peak, then by a progressive submission from several centres with different rates. For the first test, in order to simulate the case in which a node sends all the exams of the day in one bunch, 100 exams have been submitted to M5L in less than 10 minutes, filling at once all the available job slots and leaving some exams in the HTCCondor queue. For the second phase, three medical centres were configured with different parameters: a large centre submitting an exam every 10 minutes for a total of 100 exams; a medium centre submitting an exam every 20 minutes for a total of 30 exams; a small centre submitting an exam every 30 minutes for a total of 10 exams. The numbers of exams for each centre is comparable with the actual activity of a radiology department in a hospital of large, medium or small size.

2.4 Clinical Validation

In the literature there are some works aiming at evaluating the impact of a CAD system on clinical data (Beyer, 2007; Mang, 2007). These works

confirm the positive contribution of CAD as *second reader*. However all the results were obtained using retrospective studies on relative small databases. In (Beyer, 2007), for example, the author claims that possible *memory effects* could not be neglected. In addition, the majority of the studies have evaluated CADs using datasets coming from screening campaigns. Conversely, lung cancer screening campaigns do not represent the primary source of chest CT scans acquired in a hospital. In fact, clinical studies usually investigate the appearance of a pulmonary nodule as first sign of an extra-thoracic metastatic tumour. This diagnostic analysis is usually performed in oncological patients with an extra-thoracic cancer. In fact, the early detection of pulmonary metastases (combined with an adequate follow-up) can really improve the survival of the patients (Beattie, 1975). We set up a prospective observer study aiming at investigating the impact of the CAD as second-reader in the detection of pulmonary metastases. The second aim of the study was to clinically validate the M5L on-demand system, since we believe that only a direct usage in clinical practice could give us the possibility to test all the functionalities of the developed system, and build new ones if requested by clinicians. Finally, this observer study has the appeal to collect a dataset of annotated clinical data to be used for further clinical investigations. We set up a collaboration with the Radiology Department of the IRCCS in Candiolo, Italy. This study was a single-centre cross sectional study. Each participant underwent chest CT clinical examinations. Local institution board approval was obtained to publish the data. Two different kinds of examinations were used for the patients in our study: CT with or without contrast enhancement. Two experienced radiologists (20-35 years of experience) and one young radiologist (training as resident radiologist, 2 years of experience) took part in our study. All the radiologists work in the Radiology Department of IRCCS Candiolo. Each exam is submitted to the M5L front-end as soon as acquired, anonymized and stored in the hospital PACS. To make the procedure faster we allowed the submission of bunches of 10 cases in each upload. As soon as submitted, the three radiologists receive an e-mail with the direct link to annotate the case. Consistent with the second-reader protocol CAD results become available to radiologists only when the first unassisted reading has been completed and validated. Then, the radiologists receive the link to access and review CAD results. To speed up the second reading, CAD marks are automatically compared by our system to

the pathological ROIs in each annotation. A matching algorithm associates two findings when the 3D Euclidean distance between them is smaller than the mean diameter. In this case, the CAD finding takes the same properties (e.g. malignancy score) of the corresponding annotated finding. Using again the web-form, the radiologist can mark a finding as: False Positive, Irrelevant or True Positive. In the irrelevant findings are included the definition in (van Ginneken, 2010) and all the nodules smaller than 3 mm in diameters. Building a robust reference standard is fundamental when evaluating CAD performances (Armato, 2009). We defined a reference standard composed by all the nodules with a diameter larger than 3 mm and with a malignancy score equal to or larger than 2 (Indeterminate). The malignancy score, marked by the radiologist, is defined according to (Armato, 2011): subjective assessment of the likelihood of malignancy, assuming that the scan originated from a 60-year-old male smoker. The score goes from 0 (Highly Unlikely to be malignant) to 4 (Highly suspicious to be malignant). To build a reference standard, all the nodules have been evaluated again by a pool of two resident radiologists with about 10 years of experience, using the Osirix plugin (Section 2.2.1). They, once received the list of findings, could select a finding and scroll the images of the corresponding study. Additional measurement tools were available to assess and verify the findings size. The reading protocol was forced consensus.

3 RESULTS

3.1 Stress Test Results

After the stress test completion, a total of 240 exams were successfully analysed by M5L. On average, each exam, with about 280 slices, was processed in about 19 minutes, corresponding to the computational time of the slower algorithm (usually lungCAM) plus the time required for the combination, which is negligible. Figures 3 and 4 show a plot of the number of submitted, running and completed jobs during the first and second phase of the stress test, respectively. In the first part, the M5L CAD processed the submission of a bunch of 100 exams in about 4 hours, while during the second part the completion rate closely matched the submission rate. The gap between the start of the submissions and the peak in the number of jobs being processed, observed in both figures, corresponds to the waiting time for the re-allocation of computational resources

temporarily being used by other applications. Since the M5L CAD cluster is hosted, together with other applications, by a research infrastructure which is elastic in the management of resources, it may happen that all the resources are already allocated to other applications at the beginning of the submission. In these cases, the resources are progressively released and used to deploy M5L worker nodes. In this transition period the number of analysed exams is limited, but as soon as the full M5L computational power is accessed the completion rate is constant and consistent. Even so, in both cases the system reacted to the workload as expected and the results are fully in line with the needs for a practical clinical use of the M5L CAD. More intensive testing with a greater number of cases is planned, so as to estimate the required infrastructure size as function of the load.

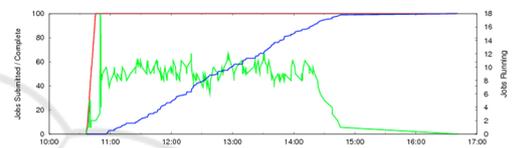


Figure 3: Distribution of jobs and completion time during the first phase of the stress-test: 100 exams submitted at once caption has more than one line so it has to be set to justify.



Figure 4: Distribution of jobs and completion time during the second phase of the stress-test: 140 exams with a slow and steady submission rate.

3.2 Clinical Validation Preliminary Results

The reference standard can be divided, according to the level of agreement, in nodules annotated by at least 1 out of 3 radiologists; nodules annotated by at least 2 out of 3 radiologists; nodules annotated by 3 out of 3 radiologists. For each category it is possible to compute the standalone CAD sensitivity, which is defined as $Sensitivity = TP / (TP + FN)$, where FN is the number of False Negatives (i.e. the nodules missed by the CAD) and TP are the nodules of the CAD matched with an annotated nodule. This quantity measures the performances of our system with respect to the reference standard. Table 1 refers to the preliminary results obtained on 70 scans. The

first row is the number of nodules inserted in the reference standard. The first three columns correspond respectively to the annotations of each radiologist. The last column refers to the list of the nodules obtained when considering the nodules in the reference standard annotated by at least three radiologists. The second row is the number of false positives computed as the fraction between all the CAD findings marked as False Positives and the total number of scans (70). The third row is the number of CAD False Negatives (FN).

Table 1: Preliminary results on 70 scans. The first three columns refer to the nodules in the reference standard annotated by each of the radiologists. The last column refers to all the nodules in the reference standard annotated by at least three RADs.

	R0	R1	R2	R0 & R1 & R2
Annotated nodules	35	38	21	17
FP/scan	3.25	4.02	3.4	3.6
FN CAD	4	6	2	2
CAD sens	90%	86%	90%	89%

CAD findings marked as True Positive and not originally included in the unassisted annotations represent nodules overlooked by the radiologists and added by the CAD. They should then be included in the reference standard, which is finally formed by all the nodules annotated by the radiologist, plus the true positives added by the CAD. Table 2 shows the preliminary results on 70 cases. The third column is the number of nodules overlooked by each radiologist, together with the relative percentage increase. The last column is the sensitivity of the radiologist plus our CAD, considered together as an only reader.

4 CONCLUSIONS

The M5L CAD was developed for the automated detection of pulmonary nodule in chest CT scans; its performance has been already extensively validated on 1043 CT scans (Lopez Torres, 2015). In order to simplify the access to the M5L CAD (or any other, in principle) across a widespread users community, a dedicated infrastructure based on a Web front-end interface and a Cloud Computing back-end has been designed, implemented and operated. On average each exam can be processed in about 19 minutes. The proposed approach solves the issue of making CAD functionality available to users without

requiring any software installation or dedicated hardware and, if properly scaled, provides the necessary amount of computational power for a large scale service. The system allows to use the CAD both as concurrent reader and second-reader mode. In this last scenario, the second reading time is sensitively decreased with an automated matching algorithm comparing annotated nodules with CAD marks: this automated procedure allows radiologists to save time and focus on overlooked findings. Preliminary results on the ongoing clinical validation show that the M5L CAD adds about a 20% of the nodules originally overlooked by the radiologists, producing a remarkable increase of the overall (RAD+CAD) sensitivity. We are planning to finalize these results on a database of 220 scans and to perform additional large scale stress tests, together with expanding the network of our users.

Table 2: Results of the second reading on 70 scans. The first three columns refer to the nodules in the reference standard annotated by each radiologist, plus CAD TPs originally overlooked by the radiologists.

	R0 & CAD	R1 & CAD	R2 & CAD	R0 & R1 & R2 & CAD
Annotated nodules	38	41	40	36
FP/scan	3.25	4.02	3.4	3.6
FN CAD	4	6	2	2
TP added by CAD (incr %)	3 (+9%)	3 (+8%)	19 (+90%)	19 (+93%)
Sensitivity	91%	87%	95%	95%

ACKNOWLEDGEMENTS

The authors thank the technical staff of the INFN Computer Centre in Torino, for their contribution in keeping the infrastructure functional at all times.

REFERENCES

Armato, S. G., Roberts, R. Y., Kocherginsky, M., et al. (2009). Assessment of radiologist performance in the detection of lung nodules: dependence on the definition of truth. *Academic radiology*, 16(1):28–38.

Armato III, S. et al. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical Physics*,

- 38(2):915–931.
- Bagnasco, S., Berzano, D., Brunetti, R., et al. (2014). Integrating multiple computing needs via a private cloud infrastructure. *Journal of Physics: Conference Series*, 513(032100).
- Beattie, E. J., Martini, N., Rosen, G., et al. (1975). The management of pulmonary metastases in children with osteogenic sarcoma with surgical resection combined with chemotherapy. *Cancer*, 35(3):618–621.
- Bellotti, R., Cerello, P., Tangaro, et al. (2007). Distributed medical images analysis on a grid infrastructure. *Future Generation Computer Systems*, 23(3):475–484.
- Berzano, D. (2014). github.com/dberzano/elastic.
- Beyer, F., Zierott, L., Fallenberg, E., et al. (2007). Comparison of sensitivity and reading time for the use of computer-aided detection (CAD) of pulmonary nodules at MDCT as concurrent or second reader. *European radiology*, 17(11):2941–2947.
- Brochu, B., Beigelman-Aubry, C., Goldmard, J., et al. (2007). Evaluation de l'impact d'un système CAD sur la performance des radiologues pour la détection des nodules pulmonaires sur des examens scanographiques multicoups du thorax. *Journal de Radiologie*, 88(4):573–578.
- Brown, M. S., Goldin, J. G., Rogers, S., et al. (2005). Computer-aided lung nodule detection in ct: Results of large-scale observer test1. *Academic radiology*, 12(6):681–686.
- Buncic, P., Aguado Sanchez, C., Blomer, J., et al. (2010). Cernvm—a virtual software appliance for LHC applications. In *Journal of Physics: Conference Series*, volume 219, page 042003. IOP Publishing.
- Camarlinghi, N., Gori, I., Retico, A., et al., (2012). Combination of computer-aided detection algorithms for automatic lung nodule identification. *Int J CARS*, 7:455–464.
- Cerello, P., Cheran, S. C., Bagagli, F., et al. (2008). The channeler ant model: object segmentation with virtual ant colonies. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 3147–3152. IEEE.
- CERNVM (2015). <http://cernvm.cern.ch/portal/ucernvm>.
- Coombs, K. (2009). Drupal done right. *Library journal*, 134(19):30–32.
- Das, M., Muhlenbruch, G., Mahnken, A., et al. (2006). Small pulmonary nodules: Effect of two computer-aided detection systems on radiologist performance 1. *Radiology*, 241(2):564–571.
- De Nunzio, G., Tommasi, E., Agrusti, et al. (2011). Automatic lung segmentation in ct images with accurate handling of the hilar region. *Journal of digital imaging*, 24(1):11–27.
- Goo, J. (2015). Computer-aided nodule detection and volumetry: role in lung cancer screening. *European Congress of Radiology*, Wien 4-8 March 2015.
- Li, Q., Sone, S., and Doi, K. (2003). Selective enhancement filters for nodules, vessels, and airway walls in twoand three-dimensional CT scans. *Medical Physics*, 30(8):2040–2051.
- Lopez Torres, E., Fiorina, E., Pennazio, et al. (2015). Large scale validation of the M5L lung CAD on heterogeneous CT datasets. *Medical Physics*, 42(4):1477–1489.
- Malvezzi, M., Bertuccio, P., Rosso, T., et al. (2015). European cancer mortality predictions for the year 2015: does lung cancer have the highest death rate in EU women? *Annals of Oncology*, 26(4):779–786.
- Mang, T., Peloschek, P., Plank, C., et al. (2007). Effect of computer-aided detection as a second reader in multidetector-row ct colonography. *European radiology*, 17(10):2598–2607.
- Matsumoto, S., Ohno, Y., Yamagata, H., et al. (2008). Computer-aided detection of lung nodules on multidetector row computed tomography using three-dimensional analysis of nodule candidates and their surroundings. *Radiation Medicine*, 26(9):562–569.
- Mell, P. and Grance, T. (2011). The NIST definition of cloud computing. Milojicic, D., Llorente, I. M., and Montero, R. (2011). Opennebula: A cloud management tool. *IEEE Internet Computing*, (2):11–14.
- Moyer, V. (2014). Screening for lung cancer: Us preventive services task force recommendation statement. *Annals of Internal Medicine*, 160(5):330–338.
- National Lung Screening Trial (2011). Reduced lungcancer mortality with low-dose computed tomographic screening. *The New England Journal of Medicine*, 365(5):395.
- Retico, A., Delogu, P., Fantacci, et al. (2008). Lung nodule detection in low-dose and thin-slice computed tomography. *Computers in biology and medicine*, 38(4):525–534.
- Retico, A., Fantacci, M., Gori, I., et al. (2009). Pleural nodule identification in low-dose and thin-slice lung computed tomography. *Computers in Biology and Medicine*, 39(12):1137–1144.
- Rosset, A., Spadola, L., and Ratib, O. (2004). OsiriX: an open-source software for navigating in multidimensional DICOM images. *Journal of Digital Imaging*, 17(3):205–216.
- Society, A. C. (2015). *Cancer facts & figures*. The Society.
- Tannenbaum, T., Wright, D., Miller, K., et al. (2001). Condor: a distributed job scheduler. In *Beowulf cluster computing with Linux*, pages 307–350. MIT press.
- Turner IV, W. P., PE, J., Seader, P., and Brill, K. (2006). Tier classification define site infrastructure performance. *Uptime Institute*, 17.
- van Ginneken, B., Armato, S. G., de Hoop, B., et al. (2010). Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study. *Medical image analysis*, 14(6):707–722.