Computer-aided detection systems to improve lung cancer early diagnosis: state-of-the-art and challenges

2017 J. Phys.: Conf. Ser. 841 012013

(http://iopscience.iop.org/1742-6596/841/1/012013)

IOP Conf. Series: Journal of Physics: Conf. Series **841** (2017) 012013 doi:10.1088/1742-6596/841/1/012013

Computer-aided detection systems to improve lung cancer early diagnosis: state-of-the-art and challenges

A Traverso[1,2], E Lopez Torres[2], M E Fantacci[3] and P Cerello[2]

[1]Department of Applied Science and Technology, Polytechnic University of Turin, Turin, Italy
[2]Turin Section of INFN, Turin, Italy
[3]Pisa Section of INFN, Pisa, Italy

E-mail: alberto.traverso@polito.it

Abstract. Lung cancer is one of the most lethal types of cancer, because its early diagnosis is not good enough. In fact, the detection of pulmonary nodule, potential lung cancers, in Computed Tomography scans is a very challenging and time-consuming task for radiologists. To support radiologists, researchers have developed Computer-Aided Diagnosis (CAD) systems for the automated detection of pulmonary nodules in chest Computed Tomography scans. Despite the high level of technological developments and the proved benefits on the overall detection performance, the usage of Computer-Aided Diagnosis in clinical practice is far from being a common procedure. In this paper we investigate the causes underlying this discrepancy and present a solution to tackle it: the M5L WEB- and Cloud-based on-demand Computer-Aided Diagnosis. In addition, we prove how the combination of traditional imaging processing techniques with state-of-art advanced classification algorithms allows to build a system whose performance could be much larger than any Computer-Aided Diagnosis developed so far. This outcome opens the possibility to use the CAD as clinical decision support for radiologists.

1. Introduction

Lung cancer still represents a health issue in developed countries. In fact, it is the leading cause of cancer-related deaths both in Europe, and United States [1, 2], and it has a very low average 5-year survival rate, around 10-17%. Most of lung cancers are diagnosed in the late stages when the survival rate is very low if compared to the diagnosis in the preliminary stage (survival rate 45-50%). When lung cancer is detected in the early stage, treatment is very successful leading in most of the cases to the complete healing. However, preliminary stages of the disease are usually asymptomatic, making the early diagnosis a real clinical challenge. Relevant technological and clinical developments have been achieved in the last decades in order to improve the early diagnosis. In particular, screening lung cancer high-risk subjects (*e.g.* heavy smokers) with low dose Computed Tomography (CT) has been proved to reduce cancer mortality up to 20% in comparison to annual screening with chest radiography [3]. However, the detection of pulmonary nodules, the early manifestations of lung cancers, is a very challenging task. In fact, the identification of pathological Regions Of Interest (ROIs) in low-dose high resolution CT scans is a difficult and time consuming task for radiologists, mainly due to the high number of slices (on average 200/300 per patient) to be read. This task is even harder during a screening campaign, when millions of slices need to be analyzed as fast as possible, without missing any clinically relevant nodule. This situation represents a real

burden for radiologists and it has been called *data explosion* [4]. In addition, it has been proved [5] that radiologists' concentration decreases during the day with a substantial impact on the overall detection sensitivity. Supporting radiologists with tools for the automated detection of pulmonary nodules in chest CT scans can produce benefits in the overall detection sensitivity. For this reason, researchers have started developing Computer-Aided Diagnosis (CAD) systems. Several studies proved an increase of radiologist's detection sensitivity when supported by those systems [6, 7]. Despite these prominent results there is a big discrepancy between the level of developments achieved and the spread of CAD systems in clinical practice, which are rarely or not used in clinical routine. In this paper we highlight the main issues currently limiting the usage of CADs in clinical practice. In Sec. 2 we present M5L, our Web- and Cloud-based CAD system for the automated detection of pulmonary nodules in chest CT scans as possible solution to presented issues. In Sec. 3 we investigate state-of-the-art CAD systems presenting the results of the medical imaging challenge we organized to investigate new imaging processing techniques to improve the overall detection sensitivity.

2. The M5L on-demand lung CAD

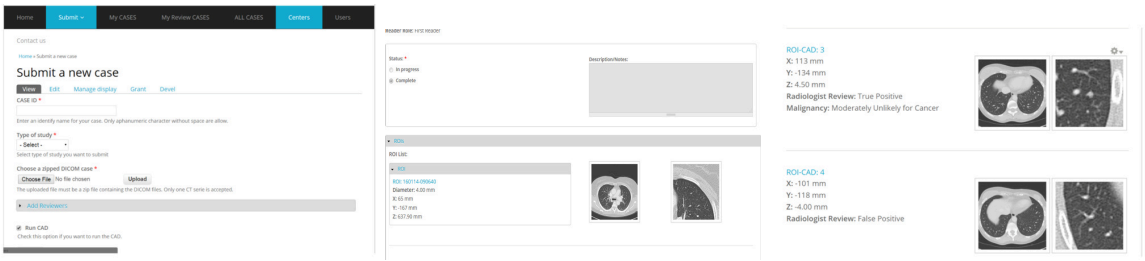
2.1. Motivation

The standard approach to make CAD algorithms available in clinical routine of health facilities consists of deploying stand-alone workstations, together with a vendor-dependent Graphic User Interface (GUI). However, this solution presents several drawbacks: high fixed costs of the software licenses and the dedicated hardware and the rapid obsolesce of both. The software is usually protected by copyright and therefore it is difficult to access the real performances of the algorithms because no scientific publications are available. Additional costs could also be charged for the integration of mentioned workstations with, for example, the PACS (Picture Archiving Communication Software) already present in the hospital. Furthermore, the computational power required increases as a function of algorithm's complexity, often requiring very expensive hardware to be bought. Finally, if a CAD system has to be used during a massive screening campaign, where thousands of scans need to be analyzed as fast a possible, the flexibility in the computation resources becomes a mandatory requirement. Since screening campaigns are usually concentrated within a limited period and considering that the number of CTs which need to be analyzed is much larger than the workload in clinical routine, buying new powerful hardware for screening would not represent a cost-effective solution. For this reason, there have been some attempts to handle the analysis of medical images on distributed environments based on a GRID infrastructure [8]. Conversely, GRID computing is not suitable for most of the medical physics projects due to its rigidity and complexity of the infrastructure. Furthermore, the management of a GRID infrastructure requires dedicated man power for the service, which is usually not available and not cost-effective in a clinical structure. Finally, guaranteeing the possibility to share CAD results and more in general medical annotations between radiologists can improve the effectiveness of the diagnosis [9]. Unfortunately, the usage of stand-alone workstations makes almost impossible the possibility to allow multipanel annotations. In addition, small clinical facilities, which usually do not have a pool of resident expert radiologists could benefit from on-line consultation from radiologists of other institutions. The M5L collaboration tried to tackle all the previous issues introducing a new paradigm of CADs in clinical practice: the M5L on-demand CAD, which does not require to the hospital any additional software or hardware installation.

2.2. Material and Methods

The M5L on-demand CAD is based on three functional blocks:

- two algorithms providing automated nodule detection;



(a) The submission module. (b) The annotation module. (c) The review module

Figure 1: Screenshots of the M5L front-end.

- a web front-end;
- a cloud-back end.

M5L is a combination of two independent CAD systems: lungCAM [10] and VBNA [11] which have been extensively validated in [12]. The web front-end has been conceived to make M5L results available to radiologists without requiring any software installation. In addition, it allows to directly insert the medical annotations of the submitted cases. The web interface can be accessed at http:/m5l.to.infn.it using any browser by desktop, tablets, and mobile devices. The system has been built using the open source content manager DRUPAL [13]. Three contributed modules have been developed for CT submission, online insertion of medical reports by radiologists (not necessarily belonging to the same institution), and CAD results access or review. The system allows to use the CAD both as *concurrent-reader* or *second-reader* [14]. To tackle the issue of making CAD algorithms available without requiring any hardware installation we set up a dedicated Cloud back-end, within the INFN Computing Center in Turin, Italy. Cloud Computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable 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 [15]. Our solution combines two basic concepts of Cloud Computing: Software as a Service (SaaS) and Infrastructure as a Service (IaaS). In fact, we provide to the users our software (M5L CAD), and give them the infrastructure (virtual machines) for CAD computations with the advantage that they do not have to take care of the maintenance.

2.3. Results

Presently, we can deploy up to 18 Virtual Machines, and a total number 48 of cores (processing units). The average processing time for each case is around 15 minutes. In order to have the possibility to scale resources, an elastic cluster was created based on CERNVM Online system [16]. Dedicated developed open-source tools [17] [18] guarantee a real-time monitoring and optimization of computing resources, so-that the user always has an up-to-date responsive system. In order to clinically validate the M5L CAD we set up a collaboration with the Institute for Cancer Research and Care (IRCCS) in Candiolo, Turin (Italy) [19]. Preliminary results show that around 20% of nodules, originally missed by radiologists during the first unassisted reading, are added by our CAD with a remarkable benefit on the overall detection sensitivity.

3. Comparing and combining algorithms: the LUNA16 challenge

A large number of CAD algorithms for the automated detection of pulmonary nodules have been discussed in the literature [20]. However, the direct comparison of the performances of those systems is not a trivial tasks. In fact, the sensitivities can vary tremendously depending on the datasets used for training and validation. A step toward an objective evaluation of

CAD systems was the ANODE09 study [21]. We organized Lung Nodule Analysis (LUNA16) an international medical challenge in order to provide an objective evaluation framework for automatic nodule detection algorithms using the largest publicly available reference database, and to investigate the combination of traditional imaging processing techniques with recent machine learning algorithms.

3.1. Material and Methods

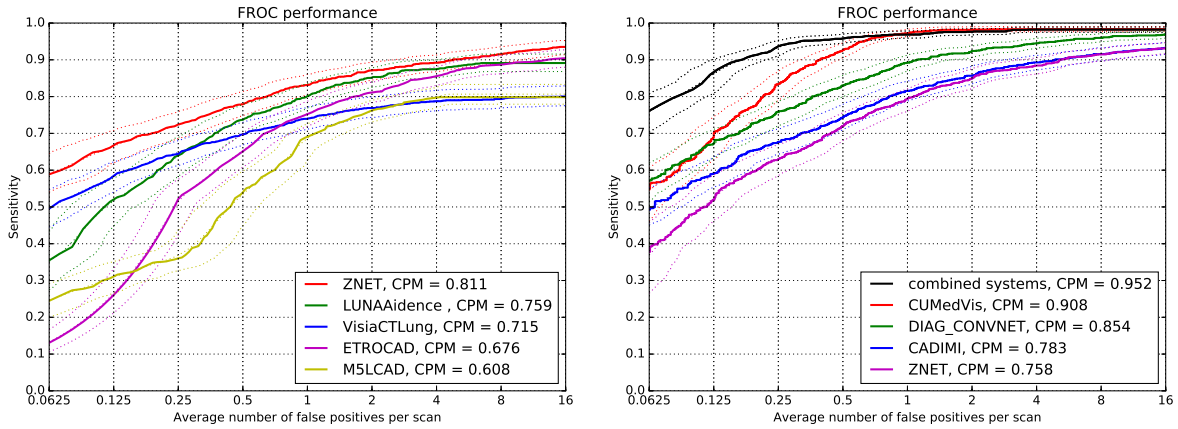
The dataset used for LUNA16 is formed by 888 scans taken from the LIDC-IDRI database [22], the biggest publicly available collection of clinical and screening annotated chest CT scans. In this database, several scanner manufacturers with different reconstruction parameters are represented, the dataset suits perfectly for CAD validation purposes. We considered as reference standard all the nodules greater than 3 mm in diameter, and annotated by at least three out of four radiologists, leading to a total of 1186 nodules. A wide range of pulmonary nodules is represented, making the automated detection a high-profile challenge. Other findings (nodules annotated by 1 or 2 radiologists, nodules smaller than 3 mm, and not nodules) were considered as irrelevant findings and not considered in the performance evaluation. We decided to divide the challenge into two tracks:

- NDET (Nodule Detection) track: participants were required to develop a complete CAD system. The input for the participants to this track are the only raw CT images;
- FPRED (False Positive Reduction) track: participants were required to propose solutions for false positive reduction stage. They were given a set of candidates, including both true nodules and false positives, and they had to provide the classification.

The FPRED track has been organized to attract machine learning groups which have not been working directly on medical imaging projects, but with the expertise on most recent machine learning techniques. Participants were able to download the full dataset, together with the list of annotations from the website of our challenge (luna16.grand-challenge.org). To avoid biases due to an open challenge (no separate training / testing datasets), they were all required to perform a 10-folds cross validation [23]. We asked participants to submit their results as a comma separated value file, which includes the list of CAD marks. For each CAD mark the position (x,y and z) coordinates, the reference to the image and the corresponding nodule probability are provided. The evaluation of true positives (TP), false positives (FP) and irrelevant findings has been performed as explained in [21]. The results are evaluated in terms of FROC (Free Response Operating Characteristic) analysis [24], where the sensitivity is plotted as a function of the average number of false positives per scan. 7 values of sensitivities at seven predefined FP rates are extracted. The overall score, called CPM (Competition Performance Metric) is then the average of previous scores, ranging from 0 (worst possible system) to 1 (best possible system).

3.2. Results

When multiple systems focus on different techniques to approach a problem, a proper combination could produce a better performance than each stand-alone system taking part in the combination. Finding a method to combine CADs without accessing their internal properties (i.e the classifier) is far from being a trivial task. Since we could only access the list of CAD marks we defined a method which makes use only of those information, similar to [21]. In addition, in order to make a fair combination, systems with high performances are weighted more than systems with a lower performance. Since the output from candidate detectors is the input for the false positive reduction systems, we combined the candidate detectors from the top performing CAD systems in NDET. We asked then to the participants to the FPRED track to run their algorithms on this combined list, and we evaluated the top performing systems



(a) (b)

Figure 2: FROC curves of the systems in (a) nodule detection track and (b) false positive reduction track. Dashed curves show the 95% confidence interval estimated using bootstrapping.

and their combination. The overall CAD is then a system which has a candidate detector, formed by a combination of different candidate detector algorithms, and a classifier, which is the combination of different false positive reduction systems. Fig. 2a shows the top performing algorithms in the NDET track. The best algorithm has a CPM score of 0.811. There are large differences in terms of sensitivity between the different systems at low false positives rates per scan. The detection of nodules with a very low FP rate is a very demanding challenge. It is worth to notice that, apart from the best system, all the other systems make use of traditional medical imaging analysis techniques. Conversely, the best system makes use of Convolutional Neural Networks (CNNs) [25] for both candidate detection, and false positive reduction. Fig. 2b shows the top performing systems in the FPRED track. The best system has a CPM score of 0.908. CNNs are used as the prediction model for all systems, showing the recent trend to apply deep learning in the medical image analysis domain. As the underlying method is similar, one could hypothesize that there could be little or no benefit when these systems are combined. On the contrary, the combination (black curve in Fig. 2b) approaches the maximum achievable sensitivity of 98.3% and saturates already starting from two false positive per scans. Although all the methods are based on CNNs, the differences in network parameters, such as the selected architectures and the input patches, makes these systems complementary in terms of the prediction.

4. Discussion

To improve the usage of CADs in clinical practice we developed M5L, a lung CAD on-demand based on the SaaS and IaaS concepts. 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. Its clinical validation on oncological patients undergoing staging or re-staging shows a positive impact of our system in the radiologists' detection sensitivity. In addition, we investigated how the combination of traditional imaging processing techniques with state-of-the-art classification algorithms could improve the overall detection sensitivity. Results show how a CAD based on this combination performs with a very high sensitivity together with a very high specificity (92% at 0.25 FP findings per scan). This solution opens the possibility to use the

CAD as *independent reader* in clinical practice as diagnosis decision support. Finally, all the combined features presented in Sec. 2 provide an innovative approach, that could be extended to other applications, allowing the possibility to combine different algorithms (as shown in Sec. 3) within the same infrastructure with a limited effort.

Acknowledgments
The authors would like to thank the technical staff of the INFN Computer Centre in Torino, for their contribution in keeping the infrastructure functional at all times, the Radiology Department of the IRCCS in Candiolo for their contribution to the clinical validation and the Diagnostic Image Analysis Group of the Radboud UMC in Nijmegen for their contribution to the organization of LUNA16.

References

[1] Malvezzi M, Bertuccio P, Rosso T, Rota M, Levi F, La Vecchia C and Negri E 2015 *Ann Oncol* **26** 779–786
[2] Siegel R L, Miller K D and Jemal A 2015 *CA: a cancer journal for clinicians* **65** 5–29
[3] Kramer B S, Berg C D, Aberle D R and Prorok P C 2011 *Journal of medical screening* **18** 109–111
[4] Rubin G D 2000 *European journal of radiology* **36** 74–80
[5] Brown M S, Goldin J G, Rogers S, Kim H J, Suh R D, McNitt-Gray M F, Shah S K, Truong D, Brown K, Sayre J W *et al.* 2005 *Academic radiology* **12** 681–686
[6] Kobayashi T, Xu X W, MacMahon H, Metz C E and Doi K 1996 *Radiology* **199** 843–848
[7] Das M, Mhlenbruch G, Mahnken A H, Flohr T G, Gndel L, Stanzel S, Kraus T, Gnther R W and Wildberger J E 2006 *Radiology* **241** 564–571
[8] Bellotti R, Cerello P, Tangaro S, Bevilacqua V, Castellano M, Mastronardi G, De Carlo F, Bagnasco S, Bottigli U, Cataldo R *et al.* 2007 *Future Generation Computer Systems* **23** 475–484
[9] Petrick N, Gallas B D, Samuelson F W, Wagner R F and Myers K J 2005 *Medical Imaging* (International Society for Optics and Photonics) pp 49–57
[10] Cerello P, Cheran S C, Bagagli F, Bagnasco S, Bellotti R, Bolanos L, Catanzariti E, De Nunzio G, Fiorina E, Gargano G *et al.* 2008 *2008 IEEE Nuclear Science Symposium Conference Record* (IEEE) pp 3147–3152
[11] Retico A, Delogu P, Fantacci M E, Gori I and Martinez A P 2008 *Computers in biology and medicine* **38** 525–534
[12] Torres E L, Fiorina E, Pennazio F, Peroni C, Saletta M, Camarlinghi N, Fantacci M and Cerello P 2015 *Medical physics* **42** 1477–1489
[13] Coombs K 2009 *Library journal* **134** 30–32
[14] Beyer F, Zierott L, Fallenberg E, Juergens K, Stoeckel J, Heindel W and Wormanns D 2007 *European radiology* **17** 2941–2947
[15] Mell P and Grance T 2011
[16] Buncic P, Sanchez C A, Blomer J, Franco L, Harutyunian A, Mato P and Yao Y 2010 *Journal of Physics: Conference Series* vol 219 (IOP Publishing) p 042003
[17] Milojičić D, Llorente I M and Montero R S 2011 *IEEE Internet Computing* 11–14
[18] Team C *HYPERLINK" http://research. cs. wisc. edu/htcondor/htc. html" http://research. cs. wisc. edu/htcondor/htc. html.[Links]*
[19] Traverso A and oth 2016 *Physica Medica: European Journal of Medical Physics* **32** 94
[20] Sluimer I, Schilham A, Prokop M and Van Ginneken B 2006 *Medical Imaging, IEEE Transactions on* **25** 385–405
[21] van Ginneken B, Armato S G, de Hoop B, van Amelsvoort-van de Vorst S, Duindam T, Niemeijer M, Murphy K, Schilham A, Retico A, Fantacci M E *et al.* 2010 *Medical image analysis* **14** 707–722
[22] Armato III S G, McLennan G, Bidaut L, McNitt-Gray M F, Meyer C R, Reeves A P, Zhao B, Aberle D R, Henschke C I, Hoffman E A *et al.* 2011 *Medical physics* **38** 915–931
[23] Efron B and Tibshirani R 1997 *Journal of the American Statistical Association* **92** 548–560
[24] Hanley J A *et al.* 1989 *Crit Rev Diagn Imaging* **29** 307–335
[25] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436–444