La radiologia medica
https://doi.org/10.1007/s11547-020-01314-8

COMPUTER APPLICATION

Radiomic features for prostate cancer grade detection through formal verification

Antonella Santone[1] · Maria Chiara Brunese[1] · Federico Donnarumma[1] · Pasquale Guerriero[1] · Francesco Mercaldo[1] · Alfonso Reginelli[2] · Vittorio Miele[3] · Andrea Giovagnoni[4] · Luca Brunese[1]

Received: 29 January 2020 / Accepted: 16 November 2020

Abstract

Aim Prostate cancer represents the most common cancer afflicting men. It may be asymptomatic at the early stage. In this paper, we propose a methodology aimed to detect the prostate cancer grade by computing non-invasive shape-based radiomic features directly from magnetic resonance images.

Materials and methods We use a freely available dataset composed by coronal magnetic resonance images belonging to 112 patients. We represent magnetic resonance slices in terms of formal model, and we exploit model checking to check whether a set of properties (formulated with the support of pathologists and radiologists) is verified on the formal model. Each property is related to a different cancer grade with the aim to cover all the cancer grade groups.

Results An average specificity equal to 0.97 and an average sensitivity equal to 1 have been obtained with our methodology.

Conclusion The experimental analysis demonstrates the effectiveness of radiomics and formal verification for Gleason grade group detection from magnetic resonance.

Keywords Formal methods · Model checking · Radiomics · Gleason grade group · Prostate

Introduction

Prostate cancer is commonly diagnosed by prostate biopsy or during trans-urethral resection for prostatic hyperplasia. The grade group describes the aggressiveness, and it is the grade that is the major determinant as to whether the patient undergoes definitive treatment or active surveillance [5, 21].

When the cancer is found, the pathologist assigns to the cancer a grade, called *Gleason Score* or *Gleason grade group*.

To assign the cancer grade, the pathologist checks the prostate tissue samples to see how much the tumour tissue is like the normal prostate tissue and to find the two main cell patterns [26]. The primary pattern describes the most common tissue pattern, and the secondary pattern describes the next most common pattern [9]. Each pattern is given a grade, with minimum grade related to the most like normal prostate tissue and the maximum one representing the most abnormal. The two grades are then added to obtain a Gleason grade groups.

In this way, a diagnosed prostate cancer can be marked with one of the following pathology-defined group: Gleason grade group *3 + 3 = 6* (*GG1*), Gleason grade group *3 + 4 = 7* (*GG2*), Gleason grade group *4 + 3 = 7* (*GG3*), Gleason grade group *4 + 4 = 8* (*GG4*) and Gleason grade group *9-10* (*GG5*).

In last years, the field of medical image analysis has attracted interest by research community [4, 18], with an increased number of pattern recognition tools and datasets freely available for research purposes [6, 33].

By analysing the state-of-the-art literature, we speculated that it can be possible to analyse medical images to obtain, through appropriate mathematical methods and algorithms, quantitative information [25] that cannot be detected through

✉ Francesco Mercaldo
francesco.mercaldo@unimol.it

[1] Department of Medicine and Health Sciences "Vincenzo Tiberio", University of Molise, Campobasso, Italy

[2] Department of Precision Medicine, University of Campania "Luigi Vanvitelli", Napoli, Italy

[3] AOU Careggi University Hospital, Firenze, Italy

[4] Department of Radiology, Ospedali Riuniti, Universit Politecnica delle Marche, Ancona, Italy

their simple visual observation by the specialist. This practice is called radiomics [2, 24].

In the state-of-the-art literature, there exist several research papers discussing the potential of radiomics to build predictive models for cancer detection, mainly exploiting artificial intelligence [19, 20, 20, 25, 35]. The main drawback related to the application of artificial intelligence in the medical context is the lack of explainability and reliability [7, 12, 30]: as a matter of fact, the model knowledge is provided by well-known algorithms, that automatically are able to output the prediction, making radiologists unconscious about the process that determined a certain decision. In contrast to artificial intelligence, we propose the use of formal verification techniques, where the domain experts (in this case the pathologists and radiologists) formulate a series of properties to be verified, therefore encapsulating the own knowledge and experience within the system predictive, making the decision of the system no longer unaware and based on an algorithm, but based on the knowledge of domain experts. Moreover, formal verification does not require a great amount of data for the property generation, differently from solutions based on artificial intelligence [5]; this avoids also the introduction of bias in training set.

For these reasons, in this paper a method to detect the prostate cancer grade group exploiting formal methods [10, 16] is presented. In particular, we exploit a set of radiomic features, obtained from magnetic resonance images (i.e. MRIs), and thus, by exploiting formal verification techniques, we label a prostate cancer MRI with the related Gleason grade group. This is resulting in a non-invasive approach from the patient's point of view (i.e. to detect the cancer grade group the biopsy is not required).

Materials and method

In this section, we first illustrate the materials used (dataset, patient population, imaging, radiomic features, statistical analysis), and then, we explain the proposed formal methodology to detect the prostate cancer grade.

Materials

First of all, we use a dataset from the Cancer Imaging Archive,[1] a large archive of tumour medical images available for research purpose. The dataset is available at the following url: https://wiki.cancerimagingarchive.net/display/Public/SPIE-AAPM-NCI+PROSTATEx+Challenges.

The dataset contains the pathologist report with the Gleason grade group details.

The prostate MR imaging was performed at the Radboud University Medical Centre (Radboudumc) in the Prostate MR Reference Center under supervision of prof. Dr. Barentsz, located in Nijmegen, The Netherlands. The dataset was collected for research in computer-aided diagnosis of prostate MR under supervision of Dr. Huisman, Radboudumc. We considered T2-weighted (T2W) images on coronal plane. T2-weighted images were acquired using a turbo spin echo sequence and had a resolution of around 0.5 mm in plane and a slice thickness of 3.6 mm. The segmentation was manually performed by exploiting the 3D Slicer software,[2] an open source software platform for medical image informatics, image processing and visualization. Moreover, for image visualization we take into account the LIFEx software,[3] a freeware software for medical images visualization.

To decide the patients to include in the study, the radiologist indicated areas of suspicion with a score per modality using a point marker. When an area was considered likely for cancer a biopsy was performed. In detail, the areas of suspicion related to each patient are the central gland (CG) and the peripheral zone (PZ) outlines marked by Drs. Nicolas Bloch (Boston University School of Medicine) and Mirabela Rusu (Case Western University) or Drs. Henkjan Huisman, Geert Litjens or Jurgen Futterer at RUNMC Netherlands. All biopsies were performed under MR guidance, i.e. in-bore MRI-guided biopsies and confirmation scans of the biopsy needle in situ were made to confirm accurate localization. Biopsy specimen was subsequently graded by a pathologist, and these results were used as ground truth.

The MRI was collected from different continents, for instance from the University of Chicago to the Harvard University. In total 15 institutions contributed to build the full dataset. Ethics committee/IRB was obtained and patient informed consent was obtained.

Below, we describe the radiomic features. To be precise, four radiomic features are considered to generate the formal model: we consider *shape features*, i.e. features independent from the grey level intensity distribution in the cancer region of interest (i.e. ROI) [34].

The details about the radiomic features are shown in Table 1.

To evaluate the method in prostate cancer Gleason grade group detection, we consider following metrics: specificity, Sensitivity, Positive Predictive Value and Negative Predictive Value.

The sensitivity of a test is the proportion of people who test positive among all those who actually have a certain Gleason grade group and it is defined as:

[1] https://wiki.cancerimagingarchive.net/.

[2] https://www.slicer.org/.

[3] https://lifexsoft.org/.

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

Table 1 The four shape radiomic features involved in the study: MinorAxisLength, MajorAxisLength, Maximum2DDiameterColumn and Maximum2DDiameterSlice

Feature	Name	Description
RF1	MinorAxisLength	This feature yields the second-largest axis length of the ellipsoid
RF2	MajoraxisLength	This feature considers the largest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component
RF3	Maximum2DDiameterColumn	It is defined as the largest pairwise Euclidean distance between tumour surface mesh vertices in the row-slice plane
RF4	Maximum2DDiameterSlice	It is defined as the largest pairwise Euclidean distance between tumour surface mesh vertices in the row-column plane

$$\text{Sensitivity} = \frac{tp}{tp + fn}$$

where *tp* indicates the number of true positives and *fn* indicates the number of false negatives

The specificity of a test is the proportion of people who test negative among all those who actually do not have that grade group and it is defined as:

$$\text{Specificity} = \frac{tn}{tn + fp}$$

where *tn* indicates the number of true negatives.

The Positive Predictive Value (PPV) is the probability that following a positive test result, that individual will truly have that specific Gleason grade group. It is defined as:

$$\text{PPV} = \frac{tp}{tp + fp}$$

The Negative Predictive Value (NPV) is the probability that following a negative test result, that individual will truly not have that specific Gleason grade group. It is defined as:

$$\text{NPV} = \frac{tn}{tn + fn}$$

where *fn* indicates the number of false negatives.

The radiomic features are obtained using a Python script developed by authors invoking pyradiomics,[4] a library for radiomic features computing from medical imaging.

Methods

Below we describe the method, discussing the formal methodology to automatically infer the Gleason grade group. In particular, we discuss how we generate the formal model and its verification through the properties generated with the knowledge of pathologists and radiologists.

Figure 1 shows the work-flow of the proposed approach for detecting prostate cancer Gleason grade group.

In a nutshell, the proposed method takes as input an MRI and it generates a formal model from the MRI slices. Thus, this model is verified with a set of properties (one property for each Gleason grade group) with the formal verification tool. If the formal verification tool output is *true*, the MRI is marked with the Gleason grade group indicated by the property. In the following, we depict the proposed approach in details.

Our analysis starts directly from MRIs. Once the patient MRI is obtained, it is possible to invoke the proposed method by gathering the MRI directly from the Picture Archiving and Communication System (*PACS* in Fig. 1). Each MRI is composed by several slices (*Slices* in Fig. 1), which are marked by the *Radiologist* to produce the *Slices segmentation*. In particular, ROI-segmented coronal slices are considered in this work. Once obtained the MRI slices and the relative segmentation, in the *Radiomic Features Extraction* step in Fig. 1, we compute the numeric values (in mm) for the *RF1*, *RF2*, *RF3* and *RF4* features from each slice belonging to the patient MRI.

The next step of the Formal Model Generation is the *Discretization*, invoked to discretize each numeric feature. We consider the method proposed by authors in [13] for the discretization. In a nutshell, we divide the features in three intervals: *low*, *basal* and *up* with the equal-width partitioning. The discretized features are converted into a formal model described in the *Language Of Temporal Ordering Specification* (LOTOS) process, a process calculus [22]. For further details, we suggest [3, 28].

To understand the way in which the formal model is generated, let us consider the example in Fig. 2.

The proposed method starts by analysing a set of segmented slices and, from each slices, the shape-based radiomic features are computed. Thus, we discretize the numeric features into three intervals (i.e. *low*, *basal* and *up*). Discretization is required for obtaining an appropriate solution. It transforms the initially continuous problem into a discrete problem. This is necessary due to the finite nature of the formal model we have to generate and to restrict the space of possible values that radiomic features can exhibit and also minimizing the impact of outliers. In particular, each line

[4] https://pyradiomics.readthedocs.io.

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

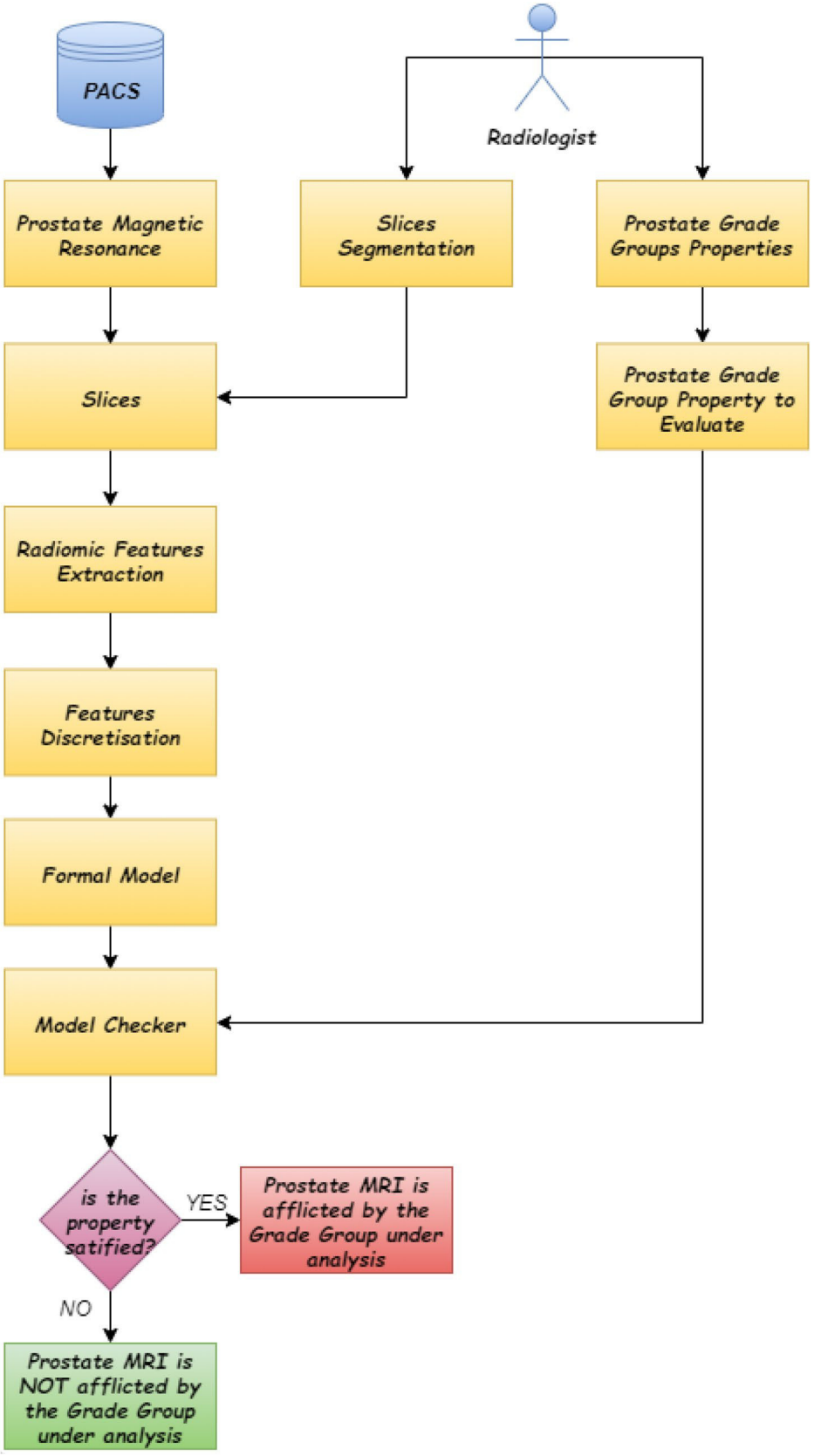

Fig. 1 The work-flow of the proposed approach: it is visible the model checker that considers as input the formal model and the prostate grade property

| Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020 |

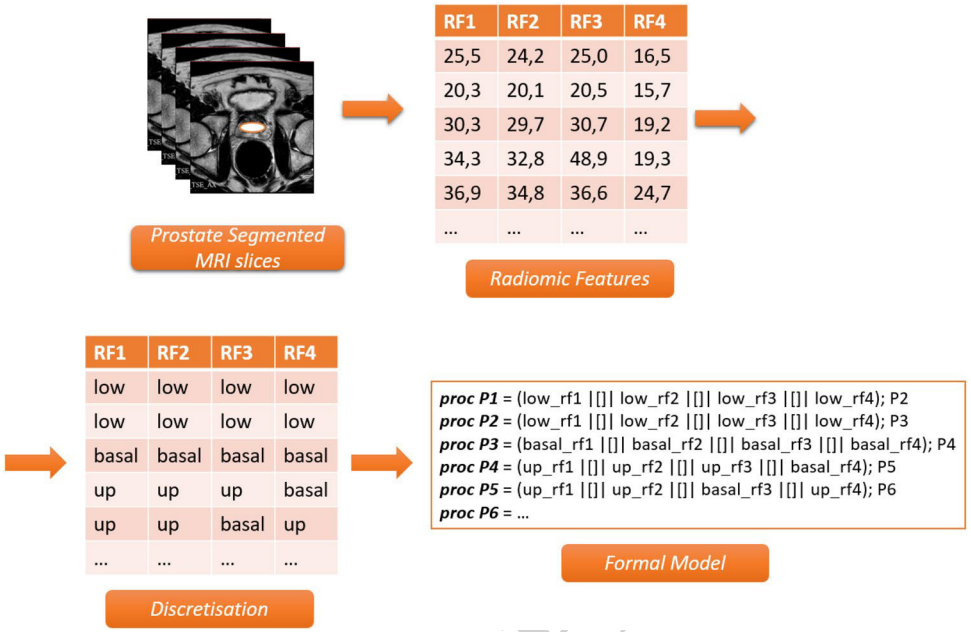

Fig. 2 Formal model generation: from the magnetic resonance slices to the radiomic features for generating the formal model

represents the discretized values of the radiomic features for a single slice, which is formalized into LOTOS processes. We have used a simplified version of the LOTOS syntax for a better readability.

In the following, a list of the used operators is reported:

- the ";" operator represents the sequentialization of actions. For example, *a*; *b* means that the event *b* must be performed after the event *a*.
- the "|[ ]|" operator represents the parallelism among events, i.e. the interleaving of the radiomic features. For example, *a* |[]| *b* means that either the event *b* and the event *a* must be performed, in any order (both the sequence *a b* and the sequence *b a*) are considered.

In the first line of the LOTOS process fragment depicted in Fig. 2 (i.e. the formal model), we observe that the *RF*1, *RF*2, *RF*3 and *RF*4 exhibit a *low* value, coherently with the discretized values previously obtained. These events are combined using the "|[ ]|" operator, while the second MRI slice, belonging to the same patient and modelled by the LOTOS process *P*2 is composed using the ";" operator. *P*2 represents the radiomic features that exhibit the same *low* values.

In the third line, the *P*3 LOTOS process in Fig. 2 codifies the radiomic features exhibit *basal* values, while the *P*4 LOTOS process codifies the fourth line, where *RF*1, *RF*2 and *RF*3 exhibit an *up* value, while *RF*4 shows a *basal* one. The LOTOS process codifies the fifth slice in a similar way.

The model has been generated to consider all points of the image, and it is analysed by a computer (not by a man) in order not to lose any details. Thus, the formal model is used to evaluate the properties formulated by pathologists and radiologists. In fact, with the support of pathologists and radiologists, for each prostate Gleason grade group, a property is formulated to detect the specific grade group and, depending on the Gleason grade group to verify, the relative property is selected, as shown in Fig. 1. We highlight that the Gleason grade group properties were assessed by expert pathologists finding confirmation of the properties effectively reflecting the grade group to detect in two different patients for each Gleason grade group.

Once the radiologist selected the property to check, the model checker is invoked: if the formal model satisfies the property, the MRI is labelled with the Gleason grade group related to the checked property; otherwise, the MRI is not related to Gleason grade group checked. The verification process is shown in detail in Fig. 3.

The model checker, coherently with the work-flow shown in Fig. 1, accepts two inputs: the formal model and the property. In Fig. 3, an example of property is depicted. The properties are expressed in μ-calculus logic [32], an extension of the propositional modal logic adding the least fixed point operator and the greatest fixed point operator. Reader unfamiliar with mu-calculus can find more information in [32]. In detail, the property is aimed to verify whether there is at least a sequence of a *low* value for *RF*1, another *low* value for *RF*1 and three *up* values for *RF*1, *RF*2 and *RF*3. Clearly, between these values there may be other values: this example of property verifies whether this sequence is present in the formal model. For this it is crucial that the property effectively reflects the Gleason grade group to detect. In practice, in the property the pathologists and radiologists formalize its knowledge. In the example in Fig. 3, the sequence in the property is

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

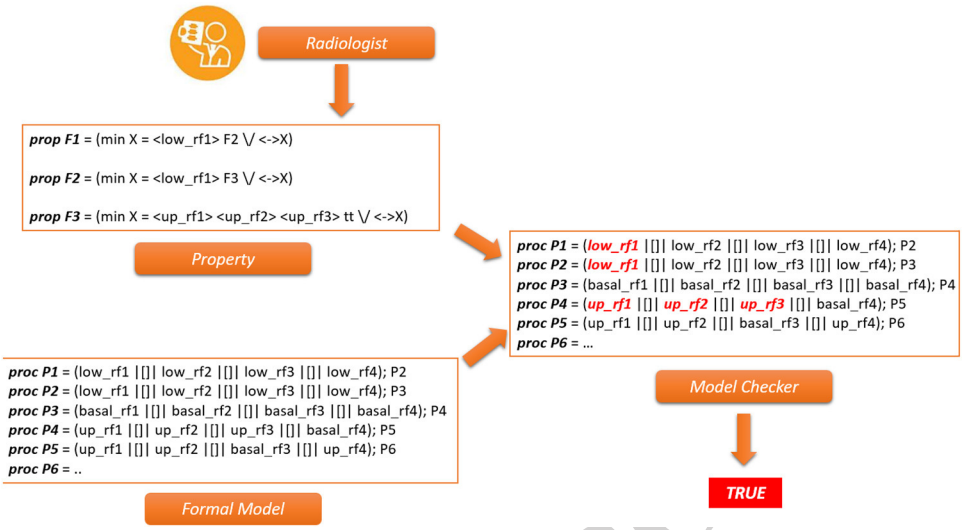

Fig. 3 The verification process: the pathologists and radiologists formulate the properties that represent the input (with the formal model) to the model checker

Table 2 Property formulated with the help of expert pathologists and radiologists for *GG5* detection

$$
\begin{aligned}
\varphi &= \mu X.\langle RF1\_basal\rangle\langle RF4\_up\rangle\varphi_2 \vee \langle -\rangle X\\
\varphi_2 &= \mu X.\langle RF2\_up\rangle\langle RF4\_up\rangle\varphi_3 \vee \langle -\rangle X\\
\varphi_3 &= \mu X.\langle RF1\_up\rangle\langle RF2\_up\rangle\langle RF4\_up\rangle\varphi_4 \vee \langle -\rangle X\\
\varphi_4 &= \mu X.\langle RF1\_up\rangle\langle RF2\_up\rangle\langle RF4\_up\rangle\varphi_5 \vee \langle -\rangle X\\
\varphi_5 &= \mu X.\langle RF1\_up\rangle\langle RF2\_up\rangle\langle RF4\_up\rangle\varphi_6\langle -\rangle X\\
\varphi_6 &= \mu X.\langle RF2\_up\rangle\langle RF3\_up\rangle\langle RF4\_up\rangle\varphi_7 \vee \langle -\rangle X\\
\varphi_7 &= \mu X.\langle RF2\_up\rangle\langle RF4\_up\rangle \mathtt{tt} \vee \langle -\rangle X
\end{aligned}
$$

present in the formal model, for this reason the model checker outputs *true*.

The pathologists and radiologists formulated five different properties: the first one aimed to detect the *GG1*, the second one to detect the *GG2*, the third one to identify the *GG3*, the forth one related to the *GG4* detection and the last one aimed to detect the *GG5*. The properties are expressed in a temporal logic to capture the variation of the cancerous area slices related to the same patient magnetic resonance.

For lack of space, we present the property describing the *GG5*, shown in Table 2. The property describes the radiomic features in the *GG5* prostate cancer: the first slice is showing a *basal* value for the radiomic feature *RF*1 and an *up* value for *RF*4 (φ in Table 2). The second slice is showing *up* value for the *RF*2 and *RF*4 features (φ_2 in Table 2). The φ_3, φ_4, φ_5 are related to *up* values for the *RF*3, *RF*4 and *RF*5 features. With regard to the φ_6, it checks whether the *RF*2, *RF*3 and *RF*4 exhibit an *up* value, while the last slice considered in the property, φ_7 is showing *up* values for the *RF*2 and the *RF*4 features. As shown by the property of the *grade group 5* prostate cancer, all the features involved show *up* values (except for a *basal* value initially shown by *RF*1 in φ): this is symptomatic that the cancer area is present and also really extended, and the continuous progression of the radiomic features to *up* values is confirming this. Properties were formulated with the assistance of expert pathologists and radiologists. In fact, pathologists and radiologists suggested, for instance, that for the *GG5*, the pattern above explained can be related to *GG5* prostate cancer. In particular, for the pathologists and radiologists, the repeated presence of the *RF*2, *RF*3 and *RF*4 with an *up* value in several slices can suggest the *GG5* grade.

The properties are checked against the patient models we have obtained from the radiomic feature set by exploiting using the Construction and Analysis of Distributed Processes [15] (*CADP*), a widespread formal verification environment providing several techniques for specifying and verifying finite-state concurrent systems. When the *CADP* formal verification environment outputs *true* when verifying a logic property on a LOTOS model, it means that the proposed method labelled the formal model as belonging to the grade group specified by the analysed property. Otherwise, the formal verification environment outputs *false*, meaning that the model under analysis is not belonging to the grade group described in the analysed formula. We recall that the properties were formulated with the help of expert pathologists and radiologists. In fact, pathologists and radiologists formulated

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

Table 3 Confusion matrix

	Total MRIs	GG1	GG2	GG3	GG4	GG5
Grade Group 1	36	**35**	0	*1*	0	0
Grade Group 2	41	*1*	**39**	*1*	0	0
Grade Group 3	20	0	0	**19**	*1*	0
Grade Group 4	8	0	0	0	**8**	0
Grade Group 5	7	0	0	0	0	**7**

In bold the patients correctly detected in the right grade group, in italic the misclassifications

the properties by looking at the discretized features for the several Gleason grade groups under analysis and they found, followed by their expertise, a common feature pattern for each Gleason grade group.

This paper represents an extension of the work proposed in [3]. We highlight below the novelties we have introduced in this work:

- we propose a method to detect the several prostate cancer grades starting from the magnetic resonance analysis. The work in [3] considers only 4 different Gleason grade groups (i.e. *3+3*, *3+4*, *4+3* and *4+4*) without considering the most aggressive Gleason grade group (*9-10*, the *GG5*) that we take in account in this paper;
- we evaluate a more extended dataset if compared with the one experimented in [3]. In fact, in [3] MRIs belonging to 60 patients were evaluated, while in this work an extended dataset of 112 patients is considered;
- in [3], we consider an algorithm to automatically infer the properties directly from a restricted set of models, while in this paper the properties are formulated by pathologists and radiologists, this is reflecting in better performances, as evidenced by the experimental analysis. This is confirming the effectiveness of the proposed properties for Gleason grade group detection aimed to formalize the knowledge of pathologists and radiologists;
- we obtain a sensitivity ranging between 0.95 and 1 and a specificity equal to 1 outperforming the performances reached in [3]. In fact, a sensitivity ranging from 0.75 to 1 and a specificity equal to 1 was obtained in [3].

We recall that the aim of the paper is to automatically detect the grade group of prostate cancer MRI. To do this, we generate a formal model from the patient MRI. Thus, a set of properties are verified, where each property is related to a different prostate cancer grade group. By invoking a formal verification environment, we check whether the properties are verified on the model: if a property is satisfied on a certain model, this model is labelled with the grade group indicated by the property.

Results

Below we present the experiment we performed to demonstrate the effectiveness of the proposed approach for prostate grade group detection

In Table 3, we show the number of MRI for each grade groups in the evaluated dataset and the number of MRI labelled as *true* by the formal verification environment.

Each row in Table 3 is related to the MRI formal model resulting *true* for the properties defined by pathologists and radiologists for each grade group. The *Total MRIs* column shown in Table 3 gives the details about MRIs involved in the experiment: a total of 112 MRIs distributed in 36 MRIs marked by pathologists and radiologists with *GG1*, 41 MRIs marked by pathologists and radiologists with *GG2*, 20 MRIs marked by pathologists and radiologists with *GG3*, 8 MRIs marked by pathologists and radiologists with *GG4* and 7 MRIs marked by pathologists and radiologists with *GG5*. In particular, the *GG1* property correctly labelled 35 MRI on 36 belonging to the *GG1*, the *GG2* property corrected labelled 39 MRI on 41, the *GG3* property correctly detected 19 MRI on 20, the *GG4* property correctly labelled all the 8 MRIs marked by pathologists and radiologists with the *GG4* disease. Finally, the *GG5* property is able to correctly detect all the 7 MRIs with the *GG5* disease belonging to the analysed dataset.

To evaluate the performance of the proposed approach following metrics we consider: Specificity, Sensitivity, Positive Predictive Value and Negative Predictive Value.

Table 4 shows the performance results.

Table 4 Performances: sensitivity ranging between 0.95 and 1 is obtained, while the specificity is equal for the analysed grade groups

	Sensitivity	Specificity	PPV	NPV
Grade Group 1	0.97	1	1	0.98
Grade Group 2	0.95	1	1	0.97
Grade Group 3	0.95	1	1	0.98
Grade Group 4	1	1	1	1
Grade Group 5	1	1	1	1

As shown by Table 4, the proposed method obtains a sensitivity ranging between 0.95 (for the *GG1* and *GG3* detection) and 1 (for the *GG4* and *GG5* detection). With regard to the specificity a value equal to 1 is reached for all the grade groups.

Discussion

In this section, we discuss and explain the choices we made in the materials and methods. We also examine the limitations of our method, and finally, we compare our results to those available from the literature to highlight our advancements.

In analysing MRIs, three planes can be considered: sagittal, coronal and axial. In designing our methodology, we have investigated the plan that led to better results. In the current literature, several papers have proved that coronal plane is the best choice, for prostate but also for other organs. In particular, for prostate lesion, in [31], the authors say that "Interestingly, models based on features extracted from T2 coronal sequence obtained much better overall performance in comparison to sagittal and transaxial sequences."

Moreover, for internal auditory canal pathology, in [1] the authors say that "Coronal T2WI better demonstrates the hypointense lesion". And still in [23], the authors write "In both oncologic and rheumatologic applications, the coronal plane is often preferred because it enables extensive coverage and straightforward investigation of the skeleton."

As stated in the previous section, our methodology provides very good results, even if it suffers of some weaknesses. The main limitations of our methodology are related to the following issues.

- Manual definition of the formulae. The logic rule-set characterizing the prostate cancer grade needs to be designed and defined. Writing the correct rules can be a rather complex task. The positive side, however, is that once the formulae have been defined, they can be used without any modification and the methodology becomes completely automatic, and it does not require any other input from the user. Nevertheless, to help the designer to write simple temporal properties, it is possible to use the user-friendly interface (UFI) developed by one of the authors in [17]. UFI has the aim of simplifying the writing of the logic properties.
- Time performances. Our methodology has been implemented in a research prototype tool whose main aim is to demonstrate the effectiveness in prostate cancer grade identification; thus, the time performances are not the core. Although the time to obtain the results are still high, the positive counterpart is the effectiveness of the results. The problem we are dealing with is of vital importance; thus, it is better to wait even a little longer times, but obtain an average sensitivity equal to 1.

Overall, we think that these limitations do not severely restrict the applicability of our method. Our method, as the experimentation demonstrated, should be considered as a good check to get a reasonable trust in the correctness of detecting prostate cancer grade. Moreover, our method has a great advantage over machine learning techniques because it does not need training cohorts. The training phase is all transferred into the experience of the pathologists and radiologists that helps in the formulation of the temporal formulae. Therefore, our data set constitutes the validation cohort.

We now review the current state-of-the-art focused on prostate cancer detection, highlighting our advancements.

Authors in [20] design an approach for the identification of prostate cancer through Bayesian networks, while authors in [11] exploit five features to demonstrate that the median of texture features is unable to discriminate between the Gleason grade groups.

Researchers in [8] train machine learning classifiers using a set of radiomic features to evaluate the classification performance of the built models.

Researchers in [35] achieve an accuracy equal to 0.85 in prostate cancer detection considering machine learning.

In reference [14], authors exploit a set of texture features to build Bayesian classifiers. They reach an accuracy equal to 88%.

Authors in [19] design a deep learning network to detect low-grade and high-grade tumours, by reaching an accuracy of 70%.

In [29], the possibility to identify prostate cancer exploiting machine learning is investigated. Authors obtain an accuracy of 83%.

Authors in [5] considers real-time verification for prostate cancer Gleason grade group detection through the UPPAAL formal verification environment [27]. Differently from the proposed work, in reference [5] the most aggressive Gleason grade (i.e. *GG5*) is not considered.

This discussion confirms the novelties of the proposed contribution formal methods based, overcoming the performances obtained by the research methods currently proposed in the literature. Furthermore, the cited works generally do not consider the different grade groups.

Conclusions

An approach model checking-based to detect the cancer grade group is proposed in this paper. We model the patient MRIs through a LOTOS model, and we evaluate the obtained models through the CADP formal verification environment. A set of properties, related to each grade group, is AQ4

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

formulated with the help of pathologists and radiologists. As future work, we plan to model patients affected by other kind of cancers. Furthermore, we will investigate whether the proposed method can be exploited in the precision medicine context, a promising research field allowing doctors to select treatments that are most likely to help patients based on a genetic understanding of their disease.

Compliance with ethical standards

Conflict of interest All authors confirm that there are not potential conflicts of interest include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations and grants or other funding. Ethics committee/IRB was obtained and patient informed consent was obtained.

Ethical standards This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abele T, Besachio D, Quigley E, Gurgel R, Shelton C, Harnsberger H, Wiggins R (2014) Diagnostic accuracy of screening MR imaging using unenhanced axial ciss and coronal t2wi for detection of small internal auditory canal lesions. Am J Neuroradiol 35(12):2366–2370
2. Brunese L, Mercaldo F, Reginelli A, Santone A (2019a) An ensemble learning approach for brain cancer detection exploiting radiomic features. Comput Methods Programs Biomed 185:105134
3. Brunese L, Mercaldo F, Reginelli A, Santone A (2019b) Formal methods for prostate cancer gleason score and treatment prediction using radiomic biomarkers. Magn Reson Imaging 66:165
4. Brunese L, Mercaldo F, Reginelli A, Santone A (2019c) Neural networks for lung cancer detection through radiomic features. In: 2019 international joint conference on neural networks (IJCNN), IEEE, pp 1–10
5. Brunese L, Mercaldo F, Reginelli A, Santone A (2019d) Prostate gleason score detection and cancer treatment through real-time formal verification. IEEE Access 7:186236–186246
6. Brunese L, Mercaldo F, Reginelli A, Santone A (2019e) Radiomic features for medical images tamper detection by equivalence checking. Procedia Comput Sci 159:1795–1802
7. Brunese L, Mercaldo F, Reginelli A, Santone A (2020) Explainable deep learning for pulmonary disease and coronavirus covid-19 detection from x-rays. Comput Methods Programs Biomed 196:105608
8. Cameron A, Khalvati F, Haider MA, Wong A (2015) Maps: a quantitative radiomics approach for prostate cancer detection. IEEE Trans Biomed Eng 63(6):1145–1156
9. Cao R, Bajgiran AM, Mirak SA, Shakeri S, Zhong X, Enzmann D, Raman S, Sung K (2019) Joint prostate cancer detection and gleason score prediction in mp-MRI via focalnet. IEEE Trans Med Imaging 38:2496
10. Ceccarelli M, Cerulo L, Santone A (2014) De novo reconstruction of gene regulatory networks from time series data, an approach based on formal methods. Methods 69(3):298–305
11. Chaddad A, Kucharczyk M, Niazi T (2018) Multimodal radiomic features for the predicting gleason score of prostate cancer. Cancers 10(8):249
12. Cimino MG, De Francesco N, Mercaldo F, Santone A, Vaglini G (2020) Model checking for malicious family detection and phylogenetic analysis in mobile environment. Comput Secur 90:101691
13. Dougherty J, Kohavi R, Sahami M (1995) Supervised and unsupervised discretization of continuous features. In: Machine learning proceedings 1995, Elsevier, pp 194–202
14. Doyle S, Madabhushi A, Feldman M, Tomaszeweski J (2006) A boosting cascade for automated detection of prostate cancer from digitized histology. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 504–511
15. Fernandez J-C, Garavel H, Kerbrat A, Mounier L, Mateescu R, Sighireanu M (1996) CADP a protocol validation and verification toolbox. In: International conference on computer aided verification, Springer, pp 437–440
16. Francesco Nd, Lettieri G, Santone A, Vaglini G (2014) Grease: a tool for efficient nonequivalence checking. ACM Trans Softw Eng Methodol 23(3):24
17. Francesco ND, Santone A, Vaglini G (2007) A user-friendly interface to specify temporal properties of concurrent systems. Inf Sci 177(1):299–311
18. Gardin I, Grégoire V, Gibon D, Kirisli H, Pasquier D, Thariat J, Vera P (2019) Radiomics: principles and radiotherapy applications. Crit Rev Oncol Hematol 138:44
19. Huang F, Ing N, Eric M, Salemi H, Lewis M, Garraway I, Gertych A, Knudsen B (2018) Abstract b094: quantitative digital image analysis and machine learning for staging of prostate cancer at diagnosis. Cancer Res 78:B094
20. Hussain L, Ahmed A, Saeed S, Rathore S, Awan IA, Shah SA, Majid A, Idris A, Awan AA (2018) Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies. Cancer Biomark 21:1–21 (Preprint)
21. Ito Y, Udo K, Vertosick EA, Sjoberg DD, Vickers AJ, Al-Ahmadie HA, Chen Y-B, Gopalan A, Sirintrapun SJ, Tickoo SK et al (2019) Clinical usefulness of prostate and tumor volume related parameters following radical prostatectomy for localized prostate cancer. J Urol 201(3):535–540
22. Langerak R (1994) Transformations and semantics for LOTOS
23. Lecouvet F (2016) Whole-body MR imaging: musculoskeletal applications. Radiology 279(2):345–365
24. Li R, Xing L, Napel S, Rubin DL (2019) Radiomics and radiogenomics: technical basis and clinical applications. Chapman and Hall/CRC, Boca Raton
25. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H (2014) Computer-aided detection of prostate cancer in MRI. IEEE Trans Med Imaging 33(5):1083–1092
26. Marshall CH, Fu W, Wang H, Baras AS, Lotan TL, Antonarakis ES (2019) Prevalence of dna repair gene mutations in localized prostate cancer according to clinical and pathologic features: association of gleason score and tumor stage. Prostate Cancer Prostatic Dis 22(1):59
27. Mercaldo F, Martinelli F, Santone A (2019) Real-time scada attack detection by means of formal methods. In: 2019 IEEE 28th international conference on enabling technologies: infrastructure for collaborative enterprises (WETICE), pp 231–236
28. Milner R (1989) Communication and concurrency. PHI Series in computer science. Prentice Hall, Upper Saddle River
29. Nguyen T.H, Sridharan S, Marcias V, Balla AK, Do MN, Popescu G (2016) Automatic gleason grading of prostate cancer using slim and machine learning. In: Quantitative phase imaging II, International Society for Optics and Photonics, vol 9718, p 97180Y
30. Parnas DL (2017) The real risks of artificial intelligence. Commun ACM 60(10):27–31
31. Sobecki P, Gora A, Zycka-Malesa D, Sklinda K, Mykhalevych I, Przelaskowski A (2017) Feature extraction optimized for prostate lesion classification. vol Part F128534, pp 22–27

Journal : Large 11547 | Article No : 1314 | Pages : 10 | MS Code : 1314 | Dispatch : 25-11-2020

32. Stirling C (1989) An introduction to modal and temporal logics for CCS. Concurrency: theory language and architecture. Springer, Berlin, pp 2–20
33. Trebeschi S, Drago S, Birkbak N, Kurilova I, Clin A, Pizzi AD, Lalezari F, Lambregts D, Rohaan M, Parmar C et al (2019) Predicting response to cancer immunotherapy using non-invasive radiomic biomarkers. Ann Oncol 30:998
34. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RG, Fillion-Robin J-C, Pieper S, Aerts HJ (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77(21):e104–e107
35. Vos PC, Hambrock T, Barenstz JO, Huisman HJ (2010) Computer-assisted analysis of peripheral zone prostate lesions using t2-weighted and dynamic contrast enhanced t1-weighted MRI. Phys Med Biol 55(6):1719

Journal : **Large 11547**	Article No : **1314**	Pages : **10**	MS Code : **1314**	Dispatch : **25-11-2020**