IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2019, accepted December 19, 2019, date of publication December 23, 2019,
date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2961754

Prostate Gleason Score Detection and Cancer Treatment Through Real-Time Formal Verification

LUCA BRUNESE[1], FRANCESCO MERCALDO[2], ALFONSO REGINELLI[3], AND ANTONELLA SANTONE[4]

[1]Department of Medicine and Health Sciences "Vincenzo Tiberio," University of Molise, 86100 Campobasso, Italy
[2]Institute for Informatics and Telematics, National Research Council of Italy (CNR), 56124 Pisa, Italy
[3]Department of Precision Medicine, University of Campania "Luigi Vanvitelli," 80138 Napoli, Italy
[4]Department of Bioscience and Territory, University of Molise, 86090 Pesche, Italy

Corresponding authors: Francesco Mercaldo (francesco.mercaldo@iit.cnr.it) and Antonella Santone (antonella.santone@unimol.it)

ABSTRACT Currently, there are 3.1 million American men affected by prostate cancer. Early detection represents the only way to safe lives. To evaluate a prostate cancer, the most widespread rank is the so-called Gleason score, based on the microscopic cancer appearance. Once assigned to the diagnosed prostate cancer its relative Gleason score, the correct therapy to be adopted must be promptly defined. To support pathologists and radiologists in timely diagnosis, in this paper we propose a method aimed to infer the Gleason score and the prostate cancer therapy exploiting formal methods. We consider a set of radiomic features directly obtained from magnetic resonance images. For this reason the proposed method is non invasive, since it does not require a biopsy. We model magnetic resonance images of patients as timed automata networks and we assign the Gleason score and the relative treatment, exploiting a set of temporal logic properties. In the experimental analysis, the properties are verified on 36 different patients, confirming the effectiveness of the proposed method with a sensitivity and a specificity equal to 1 for all the evaluated cases in Gleason score inference, and a sensitivity equal to 0.94 and a specificity equal to 1 in treatment prediction.

INDEX TERMS Model checking, formal methods, prostate, cancer, timed automata.

I. INTRODUCTION

Prostate cancer is the development of cancer in the prostate, a gland in the male reproductive system. Most prostate cancers are slow growing; however, some cancers grow relatively quickly [1]. The cancer cells can spread from the prostate to other areas of the body, particularly the bones and lymph nodes [2]. It may initially cause no symptoms. In the later stages, it can lead to difficulty urinating, blood in the urine or pain in the pelvis, back, or when urinating [3]. Other late symptoms may include feeling tired due to low levels of red blood cells.

The American Cancer Society reports that 174,650 men will be diagnosed with prostate cancer in 2019. Currently, there are nearly 3.1 million American men living with this disease, this number is roughly equal to the population of the city of Chicago. Early inference and advances in treatment represent the only way for saving lives. Early detection of prostate cancer offers the best hope for a long term, progression-free survival [4]. A recent research shows that the five year survival rate for all men with prostate cancer is nearly 100%. The relative 10 year survival rate is 98%, and 96% for 15 years [5].

Treatment for prostate cancer may involve active surveillance, radiation therapy and external-beam radiation therapy, proton therapy, high-intensity focused ultrasound, cryosurgery, hormonal therapy, chemotherapy, or some combination: all these treatments do not involve the surgeon. Obviously, another kind of treatment is the surgery i.e., the surgical removal of the prostate, the so-called prostatectomy, that is a common treatment either for early stage prostate cancer or for cancer that has failed to respond to radiation therapy [6].

The Gleason grading system is used to help the prognosis evaluation of men with prostate cancer using samples from a prostate biopsy. Together with other parameters, it is incorporated into a strategy of prostate cancer staging which predicts prognosis and helps guide therapy [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou.

A Gleason score is given to prostate cancer based upon its microscopic appearance. Cancers exhibiting a higher Gleason score are more aggressive and have a worse prognosis [8].

A total score is computed based on how cells look under a microscope, with the first half of the score based on the dominant, or most common cell morphology, and the second half based on the non-dominant cell pattern with the highest grade [9]. These two numbers are then combined to produce a total score for the cancer: the grade of the cancer, i.e., *3+3*, *3+4*, *4+3* and *4+4*, describes how much the cancer from a biopsy looks like healthy tissue (lower score) or abnormal tissue (higher score) [10].

In recent years *radiomics* is emerging as a field of medical study focused on the extraction of a large amount of quantitative features from medical images [11]. These features, called radiomic features, have the potential to uncover disease characteristics that fail to be appreciated by the naked eye. The hypothesis of radiomics is that the distinctive imaging features between disease forms may be useful for predicting prognosis and therapeutic response for various conditions, thus providing valuable information for personalized therapy [12]. Radiomics emerged from the medical field of oncology and is the most advanced in applications within that field. However, the technique can be applied to any medical study where a disease or a condition can be tomographically imaged [13].

In this paper we define the representation of magnetic resonances in terms of timed automata networks and, using the model checking technique and the radiomic features, we infer the prostate cancer Gleason score and the treatment suggested by radiologists and pathologists.

Below we itemize the novelties of the proposed method:

- we model a magnetic resonance image as a timed automata network from a set of radiomic features. At the best of authors knowledge this is the first work from this point of view, for this reason this represents the main contribution of the paper;
- we propose a set of properties, one for each Gleason score and another one to infer the treatment type. The properties are expressed in temporal logic to better characterize the evolution of the features over the time.

Moreover, we itemize the additional contributions of the paper:

- we consider a set of non invasive radiomic features to characterize Gleason score and treatment: in fact, the value related to the feature is obtained directly from magnetic resonance images and not from tissue. For this reason the biopsy is not necessary to predict the Gleason score and the treatment;
- a real-world experimental analysis is performed to evaluate the effectiveness of the proposed method. We consider a freely available data-set for research purpose to easily allow the experiment replication;
- we use model checking technique on timed automata, a technique considered to exhaustively and automatically checking if a model meets a given specification. A preliminary attempt of using model checking has been proposed by the authors in [14], where a different modelisation of magnetic resonance images is defined with respect to this work. As shown by the experimental results, the method we are proposing in this paper obtains a sensitivity equal to 1 in inferring all the Gleason scores, while the one proposed in [14] reaches a lower sensitivity. Thus, differently from [14], with our novel real-time automata based method we reach a high reliability, which is really important in contexts like the medical one, where a misleading prediction can lead to incorrect diagnoses.

The paper proceeds as follows: Section II describes the proposed method to infer prostate Gleason score and treatment, in Section III an experiment with real-world patients is performed to demonstrate the effectiveness of the proposed method, Section IV proposes an overview related to the current state-of-the-art in prostate cancer inference and, finally, in Section V conclusion and future works area drawn.

II. THE METHOD

We propose a method based on Timed Automata [15] that are well-known formalisms to model and to verify complex system and has not been explored in the clinical context. We presume the reader to be familiar with Timed Automata. For further details, we suggest [16]. The goal is two-fold: the first is the inference of the Gleason score, the second one is the surgery treatment prediction directly from MRIs considering the same radiomic feature set.

We firstly describe the considered radiomic feature set, while subsequently we present the formal model and its verification through the properties generated with the assistance of domain experts, i.e., the radiologists.

A. THE RADIOMIC FEATURE SET

Radiomic features have the potential to uncover disease characteristics that fail to be appreciated by the naked eye. The idea behind radiomics is that the distinctive imaging features between disease forms may be useful for predicting prognosis and therapeutic response for various conditions, thus providing valuable information for personalised therapy.

We investigate the effectiveness in prostate cancer Gleason Score and treatment prediction of 71 radiomic feature belonging to following five categories:

- *First Order*: this category is related to the voxel intensities distribution within the ROI i.e., the region of interest, in this study related to the areas in the MRI interested by the cancer. We consider 1 feature belonging to this category;
- *Shape*: this feature category includes descriptors of the three-dimensional size and shape of the ROI. These features are independent from the gray level intensity distribution in the ROI and are therefore only calculated on the non-derived image and mask. We consider 14 features belonging to this category;

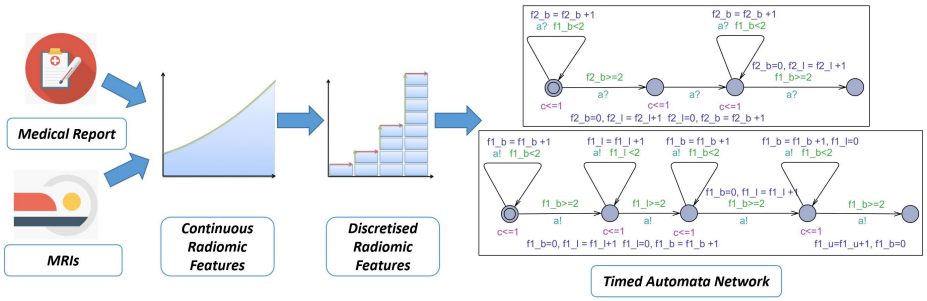

FIGURE 1. Formal model generation.

- *Gray Level Co-occurrence Matrix* (GLCM): these features consider the spatial relationship of pixels is the gray-level co-occurrence matrix i.e., the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by computing how often pairs of pixel with specific values and in a specified spatial relationship occur in an image and then extracting statistical measures from this matrix. 24 different features are belonging to this category;
- *Gray Level Run Length Matrix* (GLRLM): the grey-level run length matrix gives the size of homogeneous runs for each grey level. It quantifies gray level runs, which are defined as the length in number of pixels, of consecutive pixels that have the same gray level value. We consider 16 features belonging to this category;
- *Gray Level Size Zone Matrix* (GLSZM): the features related to this category quantify gray level zones in an image. A gray level zone is defined as the number of connected voxels that share the same gray level intensity. A voxel is considered connected if the distance is 1 according to the infinity norm. From this category, 16 different features are considered.

The full set of 71 radiomic features considered is shown in Table 5 in Appendix.

The proposed method, shown in Figures 1 and 4, consists of two phases: the *Formal Model Generation* (Figure 1) and the *Formal Model Verification* (Figure 4).

B. FORMAL MODEL GENERATION

In this section we depict the proposed timed automata based method for Gleason score and treatment detection. A timed automaton is a classical finite automaton which can manipulate clocks, evolving continuously and synchronously with the absolute time. Each transition of such an automaton is labelled by a guard, or constraint over clock values, which indicates when the transition can be fired, and a set of clocks to be reset when the transition is fired. Each location is constrained by an invariant, which restricts the possible values of the clocks for being in the state, which can then enforce a transition to be taken. For lack of space we have omitted the complete background on Timed Automata. For further details the reader can refer to [17].

The Formal Model Generation phase takes as input the set of slices belonging to the MRI related to prostate with their related Medical Report. From each slice, the numeric radiomic features are gathered (i.e., *Continuous Radiomic Features* step). The next step is the *Discretised Radiomic Features*, aimed to discretize each radiomic features. The radiomic continuous values are divided into three intervals, i.e., we map the numeric feature values into one of the following classes: *up*, *basal*, and *low*. A plethora of methods are investigated with the aim to discretize continuous values, in the current work we consider the one discussed by authors in [18]. In a nutshell Dougherty and colleagues [18] proposed to split the features in three intervals: *low*, *basal* and *up*, by considering the equal-width partitioning dividing the values of a given attribute into three equal-size intervals. The width of these intervals is calculated by exploiting the following formula: $W = (Max - Min)/3$, where Max and Min are respectively the maximum and the minimum values obtained by the feature. The equal-width partitioning has been applied to all the considered features. From the discretized features a *Timed Automata Network* is obtained. We propose a model parameterized with the full set of radiomic features, where each automaton of the network is related to one feature. Table 1 shows a fragment of discretized features.

TABLE 1. Example of feature fragmentation with three intervals i.e., u = *up*, b = *basal* and l = *low*.

Slice	F_1	F_2
1	l	u
2	b	b
3	u	b
4	u	l
5	u	b
6	b	l
7	l	b
8	b	u
9	b	u
10	l	l

The first column (i.e., Slice) indicates the slice number. We recall that each MRI belonging to a patient is composed of a set of slices. The slice number 1 is scanned before the slice 2 and so on. In order to extract meaningful information from the cancer evolution, we consider slices scanned as

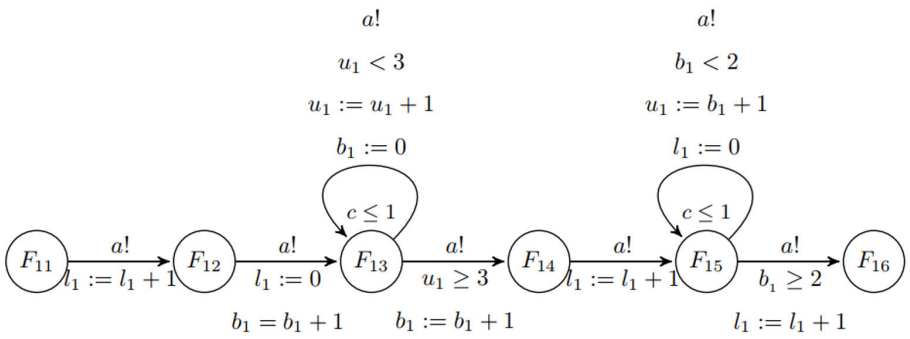

FIGURE 2. The F_1 automaton.

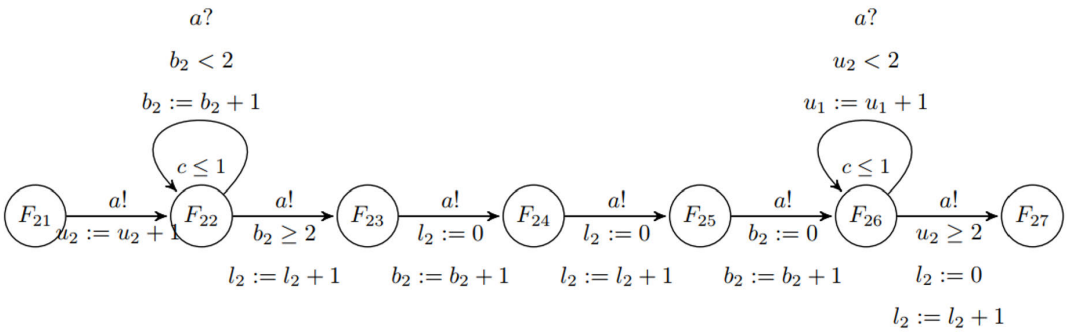

FIGURE 3. The F_2 automaton.

temporal series. The columns F_1 and F_2 indicate, respectively, the two features considered in this example.

Each patient has several slices, the intervals $u = up$, $b = basal$, $l = low$ are for this reason related to the variations that may occur between slices belonging to the same patient.

In the first line (i.e., Slice 1) of feature fragmentation depicted in Table 1, we observe that F_1 exhibits a *low* value and F_2 an *up* one.

With regard to the second slice, belonging to the same patient, we observe that F_1 exhibits a *basal* value, F_2 a *basal* one and so on.

From the discretised radiomic features the timed automata network is built: in Figures 2 and 3 are, respectively, shown the automata for the F_1 and F_2 features obtained from the discretization process in Table 1. More precisely, for each feature a timed automata is built.

In each automaton, the count of *up*, *basal* and *low* values is locally stored. In details, considering two features, $x \in \{1, 2\}$, the variables related to the F_x automaton are marked with a subscript x. For instance, in the F_1 automaton we consider as variables: u_1 for the *up* value, b_1 for the *basal* value and l_1 for the *low* one.

The channel, i.e., a in Figures 2 and 3, allowing synchronisation between automata, is not stored locally. In fact, it must guarantee the continuous and progressive automata advancement. One sender automaton, i.e., $a!$ in Figure 2, can synchronise with an arbitrary number of receivers automaton, i.e., $a?$ in Figure 3. In practice, considering that each line of the discretized features in Table 1 corresponds to the value of the features in the same time interval, the synchronization allows to switch between a time interval to the next, obliging the automata to go ahead with the next transition and to update the values of the features with the values related to the next time interval. This mechanism avoids inconsistencies between the values of the features and the time intervals. Also the clock, i.e., c in Figures 2 and 3, is not stored locally. In fact, to proceed jointly, the several automata must necessarily update their local variables at each clock cycle.

Furthermore, for each loop we note the presence of a guard, to ensure the exit from the loop. Moreover, the two automata are synchronized by using the a channel.

C. FORMAL MODEL VERIFICATION

Once generated the timed automata network and the properties, the *Formal Model Verification* (Figure 4) phase consists of checking if the properties are verified on the patient timed automata network built in the *Formal Model Generation* phase.

The *Formal Model Verification* receives as inputs the timed automata network related to one patient and a set of temporal logic properties. Specifically, five properties (*Properties* in Figure 4) are proposed: the first one, φ, to infer the *3+3* Gleason score, the second one, χ, to infer the *3+4* Gleason score, the third one, ψ, to identify the *4+3* Gleason score and

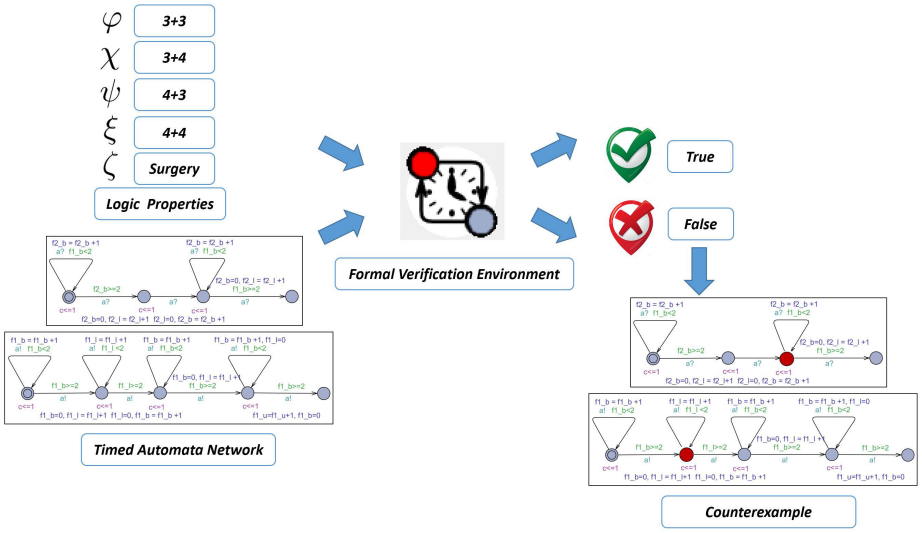

FIGURE 4. Formal model verification.

the ξ one related to the *4+4* Gleason score detection. The last properties, ζ, is aimed to predict the *Surgery* treatment.

Then, the five properties are checked against the patient timed automata network we obtained from the radiomic feature set (*Formal Verification Environment* in Figure 4) using UPPAAL, a widespread formal verification environment providing several techniques for specifying and verifying finite-state concurrent systems. When the UPPAAL formal verification environment outputs *true*, the proposed method labels the formal model as belonging to the Gleason score specified by the analysed property: *3+3* if the evaluated property is φ, *3+4* if the evaluated property is χ, *4+3* if the evaluated property is ψ or *4+4* if the evaluated property is ξ and to the *Surgery* treatment with the ζ property. Otherwise, the formal verification environment outputs *false*, meaning that the proposed method returns that the model under analysis is not belonging to the Gleason score or the surgery treatment stage described in the analysed temporal logic property. When the model checker returns *false*, there is the possibility to generate the counterexample, which is one of the most interesting advantage of the model checking, as compared to other well-known techniques for software verification. Counterexamples explain why a desired temporal logic property fails to hold. This can be of interest for the specialist to better understand the reason of the failure. The counterexamples show, in the formal model, the exact point in which the characteristics of the cancerous area are not present.

III. THE EXPERIMENT

In this section we describe the real-world data-set used to evaluate the proposed method, the logic temporal properties to infer the different Gleason scores and the treatment and, finally, the experiment we performed to demonstrate the effectiveness of the proposed formal timed automata based approach.

A. THE DATA-SET

We consider a real-world data-set belonging to the Cancer Imaging Archive,[1] a large archive of tumour medical images available for research purpose. In details, we evaluate the proposed method using a public data-set freely available for research purpose available.[2]

The data-set is composed by Prostate cancer T2-weighted coronal MRIs acquired using a 1.5 T Philips Achieva by combined surface and endorectal coil, including dynamic contrast-enhanced images obtained prior to, during and after I.V. administration of 0.1 mmol/kg body weight of Gadolinium-DTPA (pentetic acid). We considered coronal MRIs because they represent the most widespread type of image resonance always available [19], [20] in medical reports. We consider 1.5T MRI considering that across the United States and most of the world, 1.5T short-bore MRI remains the standard technology for MRI scanners,[3] this represents the most widespread technology to obtain medical images. [21].

Furthermore, the data-set contains the segmentation i.e., the ROI, pathology biopsy and excised gland tissue reports. Furthermore, it contains for each patient the report biopsy, the prostate specimen, the MRI report and the suggested treatment: these information are available at the following url: http://tiny.cc/zwladz. We consider a freely available data-set for research purpose to easily allow the experiment replication.

A total of 824 slices belonging to 36 patients are analysed. In particular, 8 patients were marked by pathologists as affected by a 3+3 Gleason prostate cancer, 12 patients were marked by pathologists as affected by a 3+4 Gleason

[1] https://wiki.cancerimagingarchive.net/
[2] https://wiki.cancerimagingarchive.net/displayPublic/PROSTATE-DIAGNOSIS
[3] https://info.atlantisworldwide.com/blog/3t-mri-vs-1.5t-mri

TABLE 2. Timed temporal logic formula for *3+3, 3+4, 4+3, 4+4* Gleason score and *surgery* detection.

$\varphi = E\langle\rangle$	$l_1 \geq 2 \wedge l_2 \geq 2 \wedge l_5 \geq 3 \wedge l_6 \geq 3 \wedge l_7 \geq 2 \wedge l_9 \geq 2 \wedge l_{11} \geq 3 \wedge l_{12} \geq 1 \wedge$ $l_{16} \geq 3 \wedge l_{19} \geq 3 \wedge l_{24} \geq 3 \wedge l_{26} \geq 3 \wedge l_{31} \geq 3 \wedge l_{42} \geq 3 \wedge l_{46} \geq 3 \wedge l_{55} \geq$ $1 \wedge l_{61} \geq 2 \wedge l_{67} \geq 2 \wedge l_{71} \geq 3$
$\chi = E\langle\rangle$	$b_5 \geq 3 \wedge b_6 \geq 3 \wedge u_8 == 1 \wedge b_7 \geq 1 \wedge b_9 \geq 2 \wedge b_{11} \geq 3 \wedge b_{14} \geq 3 \wedge b_{18}$ $\geq 2 \wedge b_{21} \geq 3 \wedge b_{26} \geq 3 \wedge b_{32} \geq 2 \wedge b_{42} \geq 3 \wedge b_{43} \geq 3 \wedge b_{46} \geq 2 \wedge b_{50}$ $\geq 1 \wedge b_{55} \geq 2 \wedge b_{57} \geq 2 \wedge b_{60} \geq 2 \wedge b_{63} \geq 2 \wedge b_{66} \geq 1 \wedge b_{70} \geq 3$
$\psi = E\langle\rangle$	$b_4 \geq 3 \wedge u_5 \geq 6 \wedge u_6 \geq 6 \wedge b_8 \geq 3 \wedge u_{11} \geq 6 \wedge b_{18} \geq 3 \wedge b_{22} \geq 2 \wedge b_{34}$ $\geq 1 \wedge b_{45} \geq 2 \wedge b_{48} \geq 2 \wedge b_{51} \geq 2 \wedge b_{55} \geq 1 \wedge b_{58} \geq 1 \wedge b_{62} \geq 2 \wedge b_{64} \geq$ $1 \wedge b_{67} \geq 1 \wedge b_{69} \geq 3 \wedge b_{70} \geq 1 \wedge b_{71} \geq 2$
$\xi = E\langle\rangle$	$u_1 \geq 4 \wedge u_2 \geq 4 \wedge u_3 == 3 \wedge u_4 \geq 4 \wedge u_{31} \geq 2 \wedge u_{32} \geq 2 \wedge u_{41} \geq 3 \wedge u_{44}$ $\geq 3 \wedge u_{47} \geq 2 \wedge u_{51} \geq 4 \wedge u_{53} \geq 4 \wedge u_{55} \geq 4 \wedge u_{56} \geq 3 \wedge u_{57} \geq 3 \wedge u_{60} \geq$ $4 \wedge u_{61} \geq 4 \wedge u_{62} \geq 1 \wedge u_{67} \geq 4 \wedge u_{69} \geq 3 \wedge u_{70} \geq 2$
$\zeta = A[\,]$	$b_{11} == 4\ imply\ A\langle\rangle\ u_6 == 3$

prostate cancer, 8 were marked by pathologists as affected by a 4+3 Gleason prostate cancer, while the remaining 8 were marked by pathologists with a Gleason Score prostate cancer equal to 4+4. Indeed, 19 patients were recommended for surgery, while the remaining 17 were not recommended for surgery. The radiologists have only marked the ROI in the MRI slices.

Figure 5 shows five different slices belonging to the evaluated data-set for four patients. In details, we show the slices number 0, 5, 10, 15 and 20 for patients exhibiting the 3+3, 3+4, 4+3 and 4+4 Gleason score prostate cancer.

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

B. THE LOGIC PROPERTIES

The Gleason score characteristics are encoded in properties expressed in a temporal: the Timed Computational Temporal Logic (TCTL) [22]. The properties aimed to make explicit the intrinsic knowledge provided by radiologists and allow to predict in which cancer stage the patient is. This is obtained using the UPPAAL formal verification tool and the formal model, previously introduced. The properties are not derived from the formal models, but are formulated by the domain experts. The formal model is created and then verified against the logic properties related to 3+3, 3+4, 4+3 and 4+4 Gleason scores and surgery treatment inference. For lack of space, for more details on the temporal logic the reader can refer to [23]. Below we present the timed temporal logic properties aimed to make explicit the intrinsic knowledge provided by radiologists.

Table 2 shows the properties. The φ property is aimed to identify the *3+3* Gleason score: this disease is present if there is a transition in the timed automata network with most of the variables with a *low* value for interval of time ≥ 2. This property highlights that the cancer area is present but it is not extended,

The χ property is related to the *3+4* Gleason score: in this case the property is satisfied when the timed automata network reaches a transition where contextually there are a series of variables with *basal* value for intervals of time ≥ 1 and the u_8 is equal to 1. Differently, from the previous formula for the *3+3* Gleason score inference, this one is aimed to verify if a medium extended cancer area is present. As suggested by domain expert, the F_8 feature would exhibit an *up* value.

The ψ property checks if a *4+3* Gleason score prostate cancer is present. As highlighted from literature [24]–[26], this Gleason score is the more difficult to recognize because is usually misclassified with the *3+4* or the *4+4* one. For this reason, machine learning based solution usually fails to classify it in the right belonging class. In this case the cancer area is extended, as expressed by the logical formula which states that a multitude of variables exhibit a *basal* value, but there are three variables exhibiting *up* values (F_5, F_6 and $F_1 1$ features).

We highlight also that the *up* values to verify the property must be present for at least 6 temporally consecutive slices, symptomatic that the area of cancer is constantly expanding.

The last Gleason score property i.e., ξ, is aimed to verify if a patient is afflicted by a *4+4* Gleason score prostate cancer. This score is inferred if a set of radiomic feature exhibit several consecutive *up* values. This is symptomatic of a really extended area where all the radiomic features are describing a rapidly increasing cancerous area.

The φ, χ, ψ and ξ Gleason scores properties are satisfied when in the formal model there exists the pattern described by the formulae. Differently, the property for the treatment detection, i.e., ζ, is in the following form: each time the $b_1 1$ variable shows a value equal to 4, the u_6 one must exhibit

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

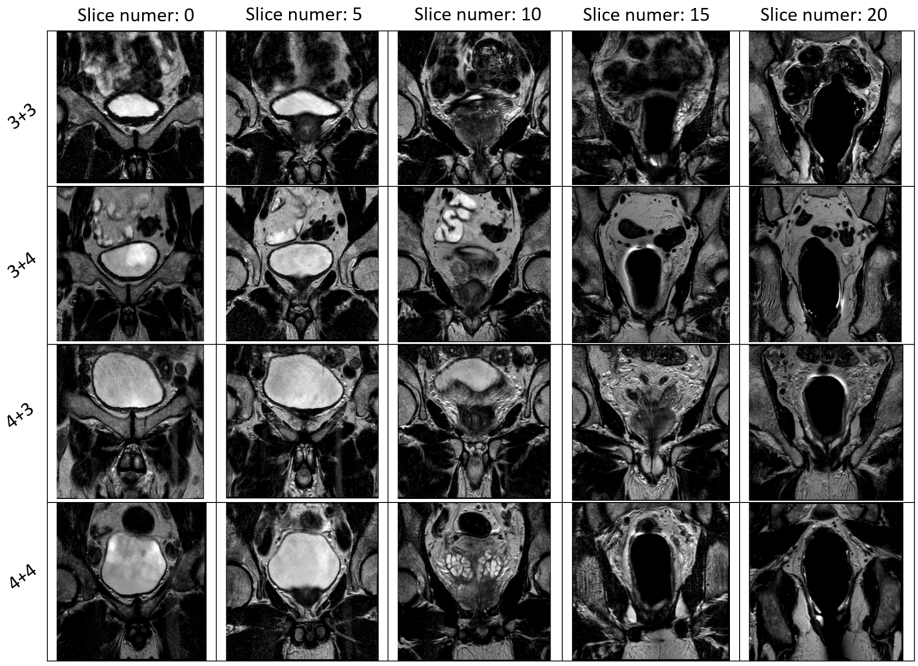

FIGURE 5. An example of MRI slices belonging to the analysed data-set.

a value equal to 3. ζ aims to verify if a prostate cancer is a plausible candidate for the *Surgery*. In Table 2 we used the UPPAAL syntax, the $\rightsquigarrow$ is replaced with the "imply" operator.

C. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we present the result we obtained from the evaluation of the proposed method. Table 3 shows the results for the formulae verification.

In Table 3 is depicted the ✓ symbol if the evaluated formula is satisfied, otherwise the ✗ symbol is added in correspondence of the model of the patient under analysis, stating that the model checker has marked the model as *false*.

Table 3 shows the model checker output for the models belonging to the 36 patients. From each patient MRI scan one timed automata network based model is obtained. The φ property, related to *3+3* Gleason score, rightly labelled all the 8 patient models affected by *3+3* Gleason prostate cancer, the χ property, related to *3+4* Gleason score, rightly labelled all the 12 patient models affected by *3+4* Gleason prostate cancer, the ψ property, related to *4+3* Gleason score, rightly labelled 8 patient models affected by *4+3* Gleason prostate cancer, while the ξ property, related to *4+4* Gleason score, rightly marked all the 8 patient models affected by *4+4* Gleason prostate cancer.

With regard to the ζ property, related to the surgery diagnosis, the model checker correctly outputs *true* on all the 18 patients marked by radiologists and pathologists as suitable candidates for the surgery, while wrongly marked the 25 patient with *true*. No formula has therefore led to erroneous results, confirming the ability of the proposed Gleason score temporal logic properties to rightly infer the Gleason score and the treatment.

Furthermore, four different metrics were used to evaluate the performance of the proposed approach: 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 the disease and it is defined as:

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

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 disease and it is defined as:

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

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 disease. It is defined as:

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

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

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

where *fn* indicates the number of false negatives.

TABLE 3. Formulae verification. For each patient we indicate the Gleason score in the *Gleason* column and the treatment. *S* stands for surgery, *NS* for no surgery in the *Treatment* column.

Formula verified on the model / Patient	Gleason	Treatment	φ	χ	ψ	ξ	ζ
1	3+3	S	✓	✗	✗	✗	✓
2	3+3	S	✓	✗	✗	✗	✓
3	3+3	S	✓	✗	✗	✗	✓
4	3+3	S	✓	✗	✗	✗	✓
5	3+3	NS	✓	✗	✗	✗	✗
6	3+3	NS	✓	✗	✗	✗	✗
7	3+3	NS	✓	✗	✗	✗	✗
8	3+3	NS	✓	✗	✗	✗	✗
9	3+4	S	✗	✓	✗	✗	✓
10	3+4	S	✗	✓	✗	✗	✓
11	3+4	S	✗	✓	✗	✗	✓
12	3+4	S	✗	✓	✗	✗	✓
13	3+4	S	✗	✓	✗	✗	✓
14	3+4	S	✗	✓	✗	✗	✓
15	3+4	S	✗	✓	✗	✗	✓
16	3+4	S	✗	✓	✗	✗	✓
17	3+4	NS	✗	✓	✗	✗	✗
18	3+4	NS	✗	✓	✗	✗	✗
19	3+4	NS	✗	✓	✗	✗	✗
20	3+4	NS	✗	✓	✗	✗	✗
21	4+3	S	✗	✗	✓	✗	✓
22	4+3	S	✗	✗	✓	✗	✓
23	4+3	NS	✗	✗	✓	✗	✗
24	4+3	NS	✗	✗	✓	✗	✗
25	4+3	NS	✗	✗	✓	✗	✓
26	4+3	NS	✗	✗	✓	✗	✗
27	4+3	NS	✗	✗	✓	✗	✗
28	4+3	NS	✗	✗	✓	✗	✗
29	4+4	S	✗	✗	✗	✓	✓
30	4+4	S	✗	✗	✗	✓	✓
31	4+4	S	✗	✗	✗	✓	✓
32	4+4	S	✗	✗	✗	✓	✓
33	4+4	S	✗	✗	✗	✓	✓
34	4+4	NS	✗	✗	✗	✓	✗
35	4+4	NS	✗	✗	✗	✓	✗
36	4+4	NS	✗	✗	✗	✓	✗

TABLE 4. Performance results.

Sensitivity	Specificity	PPV	NPV	Class
1	**1**	**1**	**1**	*3+3*
1	**1**	**1**	**1**	*3+4*
1	**1**	**1**	**1**	*4+3*
1	**1**	**1**	**1**	*4+4*
0.94	**1**	**1**	**0.96**	*Surgery*

Table 4 shows the effectiveness of the proposed method to Gleason score inference and surgery treatment stage.

The evaluation results in Table 4 show that formal methods are really promising in both Gleason score inference and surgery treatment prediction.

In fact, our method obtains following performances:

- for the *3+3* Gleason score prediction a sensitivity and a specificity equal to 1 is reached;
- for the *3+4* Gleason score prediction a sensitivity and a specificity equal to 1 is reached;
- for the *4+3* Gleason score prediction a sensitivity and a specificity equal to 1 is achieved;
- for the *4+4* Gleason score prediction a sensitivity and a specificity equal to 1 is achieved;
- for the *Surgery* prostate cancer treatment prediction a sensitivity equal to 0.94 and a specificity of 1 is obtained.

These results demonstrate the effectiveness of proposed approach formal methods based to effectively assist radiologists and pathologists in prostate cancer Gleason score inference and surgery treatment prediction by exploiting radiomic features.

Usually, machine learning solutions are more used to solve this task, as highlighted in the Related Work Section, with respect to formal methods since the latter require domain experts and logical-mathematical skills to formulate the properties. Nonetheless, we obtain very good performances since our logic temporal properties are more focused on a certain cancer than machine learning classification algorithms adopted as a black box in this context [27].

IV. RELATED WORK

In this section we review the current state-of-the-art related to prostate cancer detection. We compare the methods in literature with the one we proposed in terms of performance and proposed methodology.

Hussain and colleagues [28] propose a method to identify prostate cancer exploiting Bayesian network from MRIs i.e., by applying well-known machine learning algorithms. The main difference with our method is that the formal methods based approach proposed is able to discriminate between several Gleason Scores prostate cancers. The second is related to the detection target, authors in this work did not consider the different prostate cancer grades, while the method we propose is aimed to discriminate between several Gleason Scores prostate cancers and also aimed to predict the surgery treatment.

Chaddad and colleagues [29] consider five different features, i.e., contrast, homogeneity, difference variance, dissimilarity, and inverse difference, exploring the capacity of these features to compare between Gleason score groups with a $p < 0.05$. They demonstrate that the median of texture features is unable to discriminate between the Gleason Score groups. Differently from the method we propose, authors in this paper basically show the results of a statistical analysis, they do not propose a method to predict the Gleason score from patient MRIs. Furthermore, authors do not provide support to radiologists and pathologists with the treatment prediction.

Khalvati and colleagues [30] consider radiomic texture feature vector composed by 96 different features to detect prostate cancer in MRI images. They obtain a precision equal to 0.86 in binary labeling an MRI under analysis in tumor. The first difference with our method is related to a finer grain detection, as a matter of fact, we infer the cancer stages and the second one is related to the surgery detection. The last point is the precision obtained, that is equal to 1 for the proposed method with a number of radiomic feature equal to 4.

Authors in [31] achieve an accuracy equal to 0.85 in prostate cancer inference considering machine learning techniques. The proposed formal methods based methodology achieves a sensitivity and a specificity equal to 1 in Gleason score detection while with regard to surgery treatment

TABLE 5. Features involved in the study.

#	Class	Feature	Description
RF1	First Order	Mean	ROI average gray intensity
RF2	Shape	Elongation	relationship between two largest principal components
RF3	Shape	Flatness	relationship between largest and smallest principal components
RF4	Shape	LeastAxisLength	yield smallest axis length of the ROI-enclosing ellipsoid
RF5	Shape	MajorAxisLength	yield largest axis length of ROI-enclosing ellipsoid
RF6	Shape	Maximum2DDiameterColumn	mesh vertices in row-slice plane
RF7	Shape	Maximum2DDiameterRow	mesh vertices in the column-slice plane
RF8	Shape	Maximum2DDiameterSlice	mesh vertices in row-column plane
RF9	Shape	Maximum3DDiameter	mesh vertices
RF10	Shape	MeshVolume	volume is obtained using the surface mesh
RF11	Shape	MinorAxisLength	second-largest axis length of the ROI-enclosing ellipsoid
RF12	Shape	Sphericity	roundness of shape of the tumor region relative to a sphere
RF13	Shape	SurfaceArea	the sum of all sub-areas
RF14	Shape	SurfaceVolumeRatio	Surface Area to Volume ratio
RF15	Shape	VoxelVolume	approximate volume
RF16	GLCM	Autocorrelation	magnitude of the fineness and coarseness of texture
RF17	GLCM	ClusterProminence	skewness and asymmetry of the GLCM
RF18	GLCM	ClusterShade	skewness and uniformity of the GLCM
RF19	GLCM	ClusterTendency	voxels with similar gray-level values
RF20	GLCM	Contrast	the local intensity variation
RF21	GLCM	Correlation	linear dependency of gray level values
RF22	GLCM	DifferenceAverage	occurrences of pairs with similar and differing intensity values
RF23	GLCM	DifferenceEntropy	randomness/variability in neighborhood intensity value differences
RF24	GLCM	DifferenceVariance	heterogeneity of higher weights on differing intensity level pairs
RF25	GLCM	Id	inverse difference
RF26	GLCM	Idm	inverse difference moment
RF27	GLCM	Idmn	Inverse difference Moment Normalized
RF28	GLCM	Idn	Inverse difference Normalized
RF29	GLCM	Imc1	informational measure of correlation 1
RF30	GLCM	Imc2	informational measure of correlation 2
RF31	GLCM	InverseVariance	inverse of the variance
RF32	GLCM	JointAverage	the mean gray level intensity of the distribution
RF33	GLCM	JointEnergy	a measure of homogeneous patterns in the image
RF34	GLCM	JointEntropy	measure of the randomness/variability in neighborhood intensity values
RF35	GLCM	MCC	maximal correlation coefficient
RF36	GLCM	MaximumProbability	occurrences of the most predominant pair of neighboring intensity values
RF37	GLCM	SumAverage	occurrences of pairs with lower and higher intensity values
RF38	GLCM	SumEntropy	sum of neighborhood intensity value differences
RF39	GLCM	SumSquares	distribution of neighboring intensity level pairs
RF40	GLRLM	GLN	gray level non-uniformity
RF41	GLRLM	GLNN	gray level non-uniformity normalized
RF42	GLRLM	GLV	gray level variance
RF43	GLRLM	HGLRE	high gray level run emphasis
RF44	GLRLM	LRE	long run emphasis
RF45	GLRLM	LRHGLE	long run high gray level emphasis
RF46	GLRLM	LRLGLE	long run low gray level emphasis
RF47	GLRLM	LGLRE	low gray level run emphasis
RF48	GLRLM	RE	run entropy
RF49	GLRLM	RLN	run length non-uniformity
RF50	GLRLM	RLNN	run length non-uniformity normalized
RF51	GLRLM	RP	run percentage
RF52	GLRLM	RV	run variance
RF53	GLRLM	SRE	short run emphasis
RF54	GLRLM	SRHGLE	short run high gray level Emphasis
RF55	GLRLM	SRLGLE	short run low gray level emphasis
RF56	GLSZM	GLN	gray level non-uniformity
RF57	GLSZM	GLNN	gray level non-uniformity Normalized
RF58	GLSZM	GLV	gray level variance
RF59	GLSZM	HGLZE	high gray level zone emphasis
RF60	GLSZM	LAE	large area emphasis
RF61	GLSZM	LAHGLE	large area high gray level emphasis
RF62	GLSZM	LALGLE	large area low gray level emphasis
RF63	GLSZM	LGLZE	low gray level zone emphasis
RF64	GLSZM	SZN	size-zone non-uniformity
RF65	GLSZM	SZNN	size-zone non-uniformity normalized
RF66	GLSZM	SAE	small area emphasis
RF67	GLSZM	SAHGLE	small area high gray level emphasis
RF68	GLSZM	SALGLE	small area low gray level Emphasis
RF69	GLSZM	ZE	zone entropy
RF70	GLSZM	ZP	zone percentage
RF71	GLSZM	ZV	zone variance

detection the sensitivity is equal to 0.94 and the specificity equal to 1.

A system to assist pathologists to detect using machine learning techniques prostate cancer from MRIs of prostate histological specimens is proposed in [32]. Authors consider texture features to build Bayesian classifiers, obtaining an accuracy of 88%. Authors consider supervised machine learning classifiers, i.e., the Bayesian ones, while the proposed method obtains a specificity and a sensitivity equal to 1 overcoming the performances of well-known machine learning techniques. Furthermore, the contribution in [32] does not consider the surgery treatment prediction.

Zhang and colleagues [33] evaluate a decision tree algorithm to identify subjects with prostate cancer. They consider prostate volume, age and prostate-specific indicators as features, reaching an accuracy equal to 86.6%. Authors do not consider the different prostate cancer Gleason score where the proposed method obtained better performances, and the treatment prediction.

Huang and colleagues [34] consider a deep learning network to discriminate between low tumors and high-grade localized cases. They obtain an accuracy equal to 70%. They discriminate from low-grade and high-grade localized tumors. Differently, the method we propose is able to detect cancers between the low and the high-grade, namely the ones

labelled with the 3+4 and the 4+3 Gleason scores, obtaining better performances.

Chaddad et alius [29] consider a set of radiomic feature derived from the joint intensity matrix and the grey level co-occurrence matrix to identify the different stages of the prostate cancer. From the results point of view, using the Random Forest machine learning classification algorithm, they obtain accuracy values equal to 78.40% for Gleason score $\leq$ 3+3, 82.35% for Gleason score $=$ 3 + 4, and 64.76% for Gleason score $>$ 4 + 3. They group Gleason score MRI in three groups: the first one with the MRIs labeled with 3+3 Gleason score, the second one with the 3+4 Gleason Score, while the last group contains the 4+3 and the 4+4 Gleason score i.e., the Gleason score $>$ 4 + 3 group: this is symptomatic that when their method marks an MRI under analysis belonging to the third group the radiologist have to analyse it in order to understand if the MRI is related to the 4+3 or to the 4+4 Gleason score. Differently, we demonstrated that the proposed method is able to discriminate between the 3+3, 3+4, 4+3 and 4+4 Gleason scores, overcoming their performances in the detection of the different Gleason score, providing in addition the treatment prediction property.

Researchers in [35] investigate the possibility to identify prostate cancer exploiting the support vector machine learning algorithm. They extract a set of features from nuclei in order to build their model, obtaining an accuracy equal to 83%. Differently, the proposed method obtains a sensitivity and a specificity equal to 1 in prostate cancer stage detection. Authors in [35] do not consider the treatment prediction.

As emerged from the state-of-the-art analysis, the adoption of supervised machine learning techniques is widespread in prostate cancer detection. This confirms the novelties of our formal method based approach. The obtained results overcome the performances reached by the research methods currently proposed in literature. Moreover, the current state-of-the-art generally do not consider the different Gleason scores and the surgery treatment prediction.

V. CONCLUSION AND FUTURE WORK

Considering the significant mortality rate by prostate cancer affected patients, in this paper we propose a real-time automata approach to infer the cancer grade in terms of Gleason score and the suggested treatment to support pathologists and radiologists in early diagnosis. Differently from the researches of the state-of-the-art, which are based on the adoption of well-known machine learning techniques, our method exploits formal methods, by modeling the patient slices through a timed automata network and evaluating the obtained models through a formal verification environment. A set of properties, related to each Gleason score and the surgery treatment, is defined by the authors. The related work analysis suggests that the proposed method gains good results in the detection rate. As future work, we plan to model patients affected by other kind of cancers. Furthermore, we will investigate if the automatic property extraction can help us to make explicit more knowledge to achieve better results in Gleason score and prostate cancer treatment identification.

APPENDIX

In Table 5 there are listed the radiomic features considered in the following study.

REFERENCES

[1] L. Puca, P. J. Vlachostergios, and H. Beltran, "Neuroendocrine differentiation in prostate cancer: Emerging biology, models, and therapies," *Cold Spring Harbor Perspect. Med.*, vol. 9, no. 2, 2019, Art. no. a030593.

[2] F. La Manna, S. Karkampouna, E. Zoni, M. De Menna, J. Hensel, G. N. Thalmann, and M. Kruithof-de Julio, "Metastases in prostate cancer," *Cold Spring Harbor Perspect. Med.*, vol. 9, no. 3, 2019, Art. no. a033688.

[3] P. Paris, M. R. Cooperberg, J. Cowan, K. Lindquist, Y. Kobayashi, J. Simko, H. Bengtsson, K. L. Singh, L. F. Newcomb, and D. W. Lin, "A multibiomarker approach to predict prostate cancer pathology outcomes," Tech. Rep., 2019.

[4] D. Majera, Z. Skrott, J. Bouchal, J. Bartkova, D. Simkova, M. Gachechiladze, J. Steigerova, D. Kurfurstova, J. Gursky, and G. Korinkova, "Targeting genotoxic and proteotoxic stress-response pathways in human prostate cancer by clinically available PARP inhibitors, vorinostat and disulfiram," *Prostate*, vol. 79, no. 4, pp. 352–362, 2019.

[5] M. A. McNamara, D. J. George, K. Ramaswamy, S. Lechpammer, J. Mardekian, N. M. Schultz, L. Wang, O. Baser, A. Huang, and S. J. Freedland, "Overall survival by race in chemotherapy-Naïve metastatic castration-resistant prostate cancer (mCRPC) patients treated with abiraterone acetate or enzalutamide," Tech. Rep., 2019.

[6] M. G.-V. Ryn, D. Rosen, T. Bessede, and A. Tewari, "Optimizing outcomes during laparoscopic and robot-assisted radical prostatectomy," in *Smith's Textbook of Endourology*. 2019, pp. 1179–1193.

[7] A. G. Zhou, D. C. Salles, I. V. Samarska, and J. I. Epstein, "How are gleason scores categorized in the current literature: An analysis and comparison of articles published in 2016–2017," *Eur. Urol.*, vol. 75, no. 1, pp. 25–31, 2019.

[8] A. Surov, H. J. Meyer, and A. Wienke, "Correlations between apparent diffusion coefficient and gleason score in prostate cancer: A systematic review," *Eur. Urol. Oncol.*, to be published.

[9] C. H. Marshall, W. Fu, H. Wang, A. S. Baras, T. L. Lotan, and E. S. Antonarakis, "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 Diseases*, vol. 22, no. 1, p. 59, 2019.

[10] Y. Ito, K. Udo, E. A. Vertosick, D. D. Sjoberg, A. J. Vickers, H. A. Al-Ahmadie, Y.-B. Chen, A. Gopalan, S. J. Sirintrapun, and S. K. Tickoo, "Clinical usefulness of prostate and tumor volume related parameters following radical prostatectomy for localized prostate cancer," *J. Urol.*, vol. 201, no. 3, pp. 535–540, 2019.

[11] S. Trebeschi, S. Drago, N. Birkbak, I. Kurilova, A. M. Călin, A. D. Pizzi, F. Lalezari, D. Lambregts, M. Rohaan, and C. Parmar, "Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers," *Ann. Oncol.*, vol. 30, no. 6, pp. 998–1004, 2019.

[12] I. Tunali, J. E. Gray, J. Qi, M. Abdalah, D. K. Jeong, A. Guvenis, R. J. Gillies, and M. B. Schabath, "Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: An early report," *Lung Cancer*, vol. 129, pp. 75–79, Mar. 2019.

[13] X. Ma, L. Zhang, D. Huang, J. Lyu, M. Fang, J. Hu, Y. Zang, D. Zhang, H. Shao, and L. Ma, "Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis," *J. Magn. Reson. Imag.*, vol. 49, no. 4, pp. 1113–1121, 2019.

[14] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Formal methods for prostate cancer gleason score and treatment prediction using radiomic biomarkers," *Magn. Reson. Imag.*, to be published.

[15] K. G. Larsen, P. Pettersson, and W. Yi, "UPPAAL in a nutshell," *Int. J. Softw. Tools Technol. Transf.*, vol. 1, nos. 1–2, pp. 134–152, 1997.

[16] G. Behrmann, A. David, and K. G. Larsen, "A tutorial on UPPAAL," in *Formal Methods for the Design of Real-Time Systems*. Berlin, Germany: Springer, 2004, pp. 200–236.

[17] G. Behrmann, K. G. Larsen, O. Moller, A. David, P. Pettersson, and W. Yi, "UPPAAL—Present and future," in *Proc. 40th IEEE Conf. Decis. Control*, vol. 3, Dec. 2001, pp. 2881–2886.
[18] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretization of continuous features," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 194–202.
[19] V. Chong and Y. Fan, "Skull base erosion in nasopharyngeal carcinoma: Detection by CT and MRI," *Clin. Radiol.*, vol. 51, no. 9, pp. 625–631, 1996.
[20] M. E. Brummer, R. M. Mersereau, R. L. Eisner, and R. R. J. Lewine, "Automatic detection of brain contours in MRI data sets," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 153–166, Jun. 1993.
[21] M. Barral, P. Soyer, W. B. Hassen, E. Gayat, M. Aout, M. Chiaradia, A. Rahmouni, and A. Luciani, "Diffusion-weighted MR imaging of the normal pancreas: Reproducibility and variations of apparent diffusion coefficient measurement at 1.5-and 3.0-Tesla," *Diagnostic Interventional Imag.*, vol. 94, no. 4, pp. 418–427, 2013.
[22] E. M. Clarke, E. A. Emerson, and A. P. Sistla, "Automatic verification of finite-state concurrent systems using temporal logic specifications," *ACM Trans. Program. Lang. Syst.*, vol. 8, no. 2, pp. 244–263, 1986.
[23] R. Alur and D. L. Dill, "A theory of timed automata," *Theory Comput. Scince*, vol. 126, no. 2, pp. 183–235, 1994.
[24] Y. Yuan, W. Qin, M. Buyyounouski, B. Ibragimov, S. Hancock, B. Han, and L. Xing, "Prostate cancer classification with multiparametric MRI transfer learning model," *Med. Phys.*, vol. 46, no. 2, pp. 756–765, 2018.
[25] J. Wang, C.-J. Wu, M.-L. Bao, J. Zhang, X.-N. Wang, and Y.-D. Zhang, "Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer," *Eur. Radiol.*, vol. 27, no. 10, pp. 4082–4090, 2017.
[26] D. Albala, M. S. Manak, J. S. Varsanik, H. H. Rashid, V. Mouraviev, S. M. Zappala, E. Ette, N. Kella, K. M. Rieger-Christ, and G. R. Sant, "Clinical proof-of-concept of a novel platform utilizing biopsy-derived live single cells, phenotypic biomarkers, and machine learning toward a precision risk stratification test for prostate cancer grade groups 1 and 2 (Gleason 3+3 and 3+4)," *Urology*, vol. 124, pp. 198–206, 2019.
[27] D. L. Parnas, "The real risks of artificial intelligence," *Commun. ACM*, vol. 60, no. 10, pp. 27–31, 2017.
[28] L. Hussain, A. Ahmed, S. Saeed, S. Rathore, I. A. Awan, S. A. Shah, A. Majid, A. Idris, and A. A. Awan, "Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies," *Cancer Biomarkers*, vol. 21, no. 2, pp. 393–413, 2018.
[29] A. Chaddad, M. Kucharczyk, and T. Niazi, "Multimodal radiomic features for the predicting gleason score of prostate cancer," *Cancers*, vol. 10, no. 8, p. 249, 2018. [Online]. Available: [Online]. Available: http://www.mdpi.com/2072-6694/10/8/249
[30] F. Khalvati, A. Wong, and M. A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models," *BMC Med. Imag.*, vol. 15, no. 1, p. 27, 2015.
[31] P. C. Vos, T. Hambrock, J. O. Barenstz, and H. J. Huisman, "Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI," *Phys. Med. Biol.*, vol. 55, no. 6, p. 1719, 2010.
[32] S. Doyle, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "A boosting cascade for automated detection of prostate cancer from digitized histology," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent.* Berlin, Germany: Springer, 2006, pp. 504–511.
[33] Y. Zhang, Q. Li, Y. Xin, and W. Lv, "Differentiating prostate cancer from benign prostatic hyperplasia using PSAD based on machine learning: Single-center retrospective study in China," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 3, pp. 936–941, May 2018.
[34] F. Huang, N. Ing, M. Eric, H. Salemi, M. Lewis, I. Garraway, A. Gertych, and B. Knudsen, "Abstract B094: Quantitative digital image analysis and machine learning for staging of prostate cancer at diagnosis," *Cancer Res.*, vol. 78, no. 16, 2018.
[35] T. H. Nguyen, S. Sridharan, V. Marcias, A. K. Balla, M. N. Do, and G. Popescu, "Automatic Gleason grading of prostate cancer using SLIM and machine learning," *Proc. SPIE*, vol. 9718, Mar. 2016, Art. no. 97180Y.

LUCA BRUNESE is currently a Full Professor of imaging diagnostics. He is also the Rector of the University of Molise, Campobasso, Italy, since 2019. His research interests are mainly focused on neuroradiology and interventional radiology, where he published more than 130 contributions between journal articles and book chapters.

FRANCESCO MERCALDO received the master's degree in computer engineering from the University of Sannio, Benevento, Italy, with a thesis in software testing, and the Ph.D. degree, in 2015, with a dissertation on malware analysis using machine learning techniques. He worked as a Postdoctoral Researcher at the Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy. He is currently involved as a Lecturer in database, mobile programming (bachelor's degree) and ethical hacking (master's degree) courses at the University of Molise, Italy. His research areas include software testing, verification, and validation, with the emphasis on biomedical systems.

ALFONSO REGINELLI is currently a Researcher with the University of Campania "Luigi Vanvitelli." His research activities are related to the diagnostic imaging and radiotherapy, where he has published more than 120 conference and journal articles.

ANTONELLA SANTONE received the Laurea degree in computer science from the University of Pisa, Italy, in April 1993, and the Ph.D. degree in computer systems engineering from the Dipartimento di Ingegneria della Informazione, University of Pisa, in September 1997.

She is currently an Associate Professor of computer engineering with the University of Molise, since September 2017. She has been Assistant Professor with the University of Pisa, from November 1998 to October 2001. She has been an Associate Professor with the Department of Engineering, University of Sannio, from November 2001 to August 2017. She was involved in several research activities and projects. Her current research is focused on formal verification methods. Her research interests include formal description techniques, temporal logic, concurrent and distributed systems modeling, heuristic search, and formal methods in medical imaging. She has written more than hundred articles for international journals and conferences.