




Prostate Cancer Diagnosis Automation Using Supervised Artificial Intelligence. A Systematic Literature Review

Camilo Espinosa¹, Manuel Garcia¹, Fernando Yepes-Calderon^{2,4},
J. Gordon McComb³, and Hector Florez¹(✉) 

¹ Universidad Distrital Francisco Jose de Caldas, Bogota, Colombia
elmonachoa@gmail.com, manuel-alejo2000@hotmail.com,
haflorefz@udistrital.edu.co

² SBP LLC - RnD Department, 604, Fort Pierce, FL 34950, USA
fernando.yepes@strategicbp.net

³ Children's Hospital Los Angeles, 1300 N Vermont Ave 1006,
Los Angeles, CA 90027, USA
gmccomb@chla.usc.edu

⁴ GYM Group SA - Departamento I+R, Cra 78A No. 6-58, Cali, Colombia
fernando@gym-group.org

Abstract. Prostate cancer (PCa) is the most frequent genre-specific malignancy and the fourth overall behind lung, breast, and colon cancers. PCa is diagnosed non-invasively with serum prostate-specific antigen assay, digital rectal examination, and trans-rectal ultrasound and invasively with multiple rectal biopsies from which a Gleason score is assigned. The biopsy tissue is subject to sampling error and cellular interpretation that in turn can lead to disagreement as to whether treatment is needed, and if so, the method and the extent of therapy. Magnetic resonance (MR) imaging is proving to be progressively useful in evaluating PCa. New sequences are continually being introduced that are proving to be even more accurate in determining the extent and degree of tumor malignancy than other imaging modalities. The MR images, however, are evaluated by radiologists whose interpretation is subjective. This study reviews the currently available artificial intelligence and machine learning techniques that may eliminate the need for multiple rectal biopsies and provide a more uniform classification of these malignancies. Also, the evaluation of treatment outcome can be better assessed with more precise tumor size and classification. This paper investigates and analyzes projects related to prostate cancer's automatic diagnosis using artificial intelligence.

Keywords: Prostate cancer · MR images · Prostate automatic segmentation · Supervised classification · Machine learning · Artificial intelligence

1 Introduction

Like breast cancer for women, prostate cancer is one the most common for men, accounting for 1'276.106 new cases in 2018 and killing 358.989 men the same year [18]. Diagnosis usually starts with an elevated serum prostate (PSA) and a digital rectal examination (DRE). If the findings are of concern, the next test is transrectal ultrasound (TRUS). The process continues with the invasive transrectal biopsies performed in multiple regions of the organ to increase sampling. The extracted tissue is analyzed histologically, yielding a Gleason score (G-Score). The biopsy tissue is subject to sampling error and cellular interpretation, leading to disagreements about whether treatment is needed and the scope and method to treat. Contrary to computer tomography (CT), magnetic resonance (MR) proves to be progressively useful in evaluating prostate cancer. New sequences are continuously being introduced and have proved to be more precise in determining the extent and degree of malignancy. Additionally, MR does not use the harmful ionizing radiation present in CT and other imaging methods such as positron emission tomography (PET), digital tomosynthesis (DTS), and single positron emission computed tomography (SPECT). Currently, radiologists evaluate and derive verdicts from MR images, a process that is subject to interpretation once again.

The development of artificial intelligence (AI) has the potential to accurately and non-invasively determine the volume of the tumors, the degree of malignancy, and the response to therapy [21].

This study is a review of current AI and machine learning strategies that are proposed to evaluate MR images of prostate cancer automatically and accurately.

The document's structure is composed as follows: Sect. 2 shows method employed to develop a systematic literature review (SLR) where criteria, data extraction, quality verification, and search process have been devised. In Sect. 3, we present the investigation's results with each study's descriptions and compare accuracy and complexity.

2 Method

This study developed a Systematic Literature Review (SLR) based on the guidelines proposed by Kitchenham [11].

2.1 Research Questions

- RQ1. Which software for prostate cancer diagnosis uses artificial intelligence?
- RQ2. Which artificial intelligence techniques are implemented to classify MR images?
- RQ3. Which measurement methods exist for diagnostic in prostate cancer?

2.2 Search Process

Search Strategy. The search strategy implemented in this SLR was the Population, Intervention, and outcome strategy suggested by Kitchenham [11], which allows us to generate a suitable search chain.

- **Population.** The population in this document is defined by the software for prostate cancer diagnosis, where the keywords are `prostate cancer AND software`.
- **Intervention.** The implementation requires study of automatic segmentation and classification. Then, the keywords are: `automatic classification OR machine learning OR artificial intelligence`.
- **Outcome** The expected result is prostate cancer diagnosis with his measure. Then, the keywords are `diagnosis AND measure`.

Selected Journals and Conferences. The selected sources were found in the following digital databases: Scopus, IEEE Xplore, Springer Link, and Science Direct. The time-range in which the sources were searched was from 2015 to 2019, as shown in Table 1.

The search sentences are:

- **Scopus:**
`TITLE-ABS-KEY (("prostate cancer") AND (("artificial intelligence") OR ("machine learning") OR ("automatic segmentation")))) AND ("diagnosis")`
- **IEEE:**
`((("prostate cancer") AND (("artificial intelligence") OR ("machine learning") OR ("automatic segmentation")))) AND ("diagnosis")`
- **Springer Link:**
`ALL(("prostate cancer") AND (("artificial intelligence") OR ("machine learning") OR ("automatic segmentation"))) AND ("diagnosis")`
- **Science Direct:**
`((("prostate cancer") AND (("artificial intelligence") OR ("machine learning") OR ("automatic segmentation")))) AND ("diagnosis")`

2.3 Inclusion and Exclusion Criteria

The criteria are defined by directly related information in the case of the inclusion criteria (IC) and complementary information in the case of the exclusion criteria (EC).

Table 1. Pre-selected articles

Database	Found	Duplicates	Pre-selected
Scopus	12	2	6
IEEE Xplore	10	1	7
SpringerLink	7	0	2
ScienceDirect	5	1	3

The IC are the following:

- **IC1** Studies which refers to classifiers or automatic machines for prostate cancer diagnosis.
- **IC2** Studies in which prostate MR images are manipulated with automatic segmentation techniques.
- **IC3** Studies which refers to an automatic measurement for prostate cancer diagnosis.

The EC are the following:

- **EC1** Studies which do not focus on prostate cancer.
- **EC2** Studies in which general MR images are manipulated with automatic segmentation techniques.
- **EC3** Studies which do not use a measurement for prostate cancer diagnosis.

2.4 Data Extraction

The following data was extracted from the pre-selected articles shown in Table 2:

- Type of the article (journal or conference) with its reference.
- Author’s data with their affiliations.
- Date of the document.
- Database where the article is located.
- DOI (Digital Object Identifier).
- Abstract of the article with its topic area.
- Keywords of the article.

Table 2. Selected articles

Database	Pre-selected	Selected
Scopus	6	3
IEEE Xplore	7	3
SpringerLink	2	1
ScienceDirect	3	1

2.5 Quality Verification

The studies exposed in Table 2 had to pass through the following verification criteria (VC) for an optimal study with the collected information:

- **VC1** Does the article contain information that can answer the research questions?
- **VC2** Does the article include a good explanation about its content?
- **VC3** Does the study deal about the research?
- **VC4** Was the article published in journals or conferences?
- **VC5** Does the article have a good accuracy in its results?

Finally, the selected journals and conferences articles are show in Table 3.

Table 3. Selected journals and conferences

Name	SJR H index
Proceedings of the National Academy of Science	699
Journal of biomedical optics	123
Computerized Medical Imaging and Graphics	63
BioSMART	2
European Radiology	134
IEEE Transactions on Medical Imaging	195
2015 IEEE International Conference on Image Processing (ICIP)	97
Journal of Magnetic Resonance Imaging	142

3 Results

Prostate-cancer diagnosis based on artificial intelligence (AI) is a logical path of development. The AI's goal is to find cancerous areas on prostate images – often MRI – improving the accuracy of diagnosis. Several authors have pursued automation employing all sorts of image modalities and AI classifiers.

Fehr *et al.*, [7] proposed an automatic Gleason score (G-score) estimation using MR T2 weighted image (T2W) and apparent diffusion coefficient (ADC) maps for the transitional zone (TZ) and peripheral zone (PZ). The images require pre-processing and registration. In Fehr's outline, a support vector machine (SVM) uses texture features extracted from mp-MRI intensity and the gray level co-occurrence matrix (GLCM), consisting of energy, entropy, correlation, homogeneity, and contrast to yield a G-score appraisal. Also, the system has an accuracy of 87%, the sensitivity of 87%, and specificity of 84%.

Nguyen *et al.*, [14] implemented a random forest (RF) algorithm to classify light interference microscopy (SLIM) images from prostate biopsy to produce a G-score number. The RF was trained with a tissue microarray (TMA) composed

by H&E biopsies and a feature vector histogram for determining which pixel in the SLIM is a lumen, gland, or stroma with an accuracy of 0.82 measurements by the receiver operating characteristic (ROC) curve with its area under the curve (AUC).

Giannini *et al.*, [8] introduced a method to explore the prostate's PZ exclusively. The strategy uses T2W images with CAD (Computer-Aided Detection) to indicate candidate cancer zones. Additionally, Giannini's method requires image registration in order to standardize them. Prostate segmentation was defined by identifying a rectangular region by each slice of the mp-MRI, then the rectangle was segmented by the Hough transformation. Feature extraction develops the intensity of the ADC maps and T2W images, while pharmacokinetics act as an SVM classifier to discriminate between normal and tumor voxels. Later a false-positive (segmented tissues that are not cancerous) reduction are made to exclude them; the system has an accuracy of 91%.

Peyret *et al.*, [15] proposed an algorithm of Multispectral Linear Binary Pattern (MLBP), this handles Multispectral images, where the image features are obtained, and the texture is analyzed to get intensity. The images are next divided into blocks, with each block having feature vectors. Each block makes a codebook that is classified with SVM. This method reports an accuracy of 93.7%.

Wang *et al.* [22] developed an AI mechanism that mimics the Prostate Imaging Reporting and Data System (PI-RADS v2). This procedure uses T1-weighted imaging (T1WI), T2WI, diffusion-weighted imaging (DWI), as well as dynamic contrast-enhanced (DCE) imaging. Wang analyzes the images with a radial basis function (RBF) and determines verdicts with an SVM classifier.

A study called Focalnet is described in Cao *et al.*, [2] in which the diagnosis and lesion detection is made using a pre-registered mp-MRI for prostate cancer with a convolutional neural network (CNN) using Gleason the to characterizes the aggressiveness of the tumor. Focalnet also uses a mutual finding loss (MFL), which allows identifying the optimal features in a T2W and ADC images for the CNN training phase. The T2W image is used for assessing intensity variation. Thus, Focalnet has an accuracy of 80.5% and a sensitivity of 79.2%.

In another study, Reda *et al.*, [17] implemented a CAD system for DW-MRI to find the benign and malignant tissue in the prostate. They employed a non-negative matrix factorization (NMF) to segment the prostate with DW-MRI. Also, the ADC values are calculated for feature extraction, as refined by a Gauss-Markov Random Field (GGMRF). The cumulative distribution function (CDF) universalized the benign and the malignant features extracted to train a stacked nonnegativity constraint autoencoder (SNCAE). The CAD system has an accuracy of 100% in a dataset consisting of 53 cases.

Ginsburg *et al.* proposes another CAD system [9], which arguments mp-MR images with a Gleason score to obtain a diagnosis. The features were extracted using the intensities of the MRI measure of concordance using an intra-class correlation coefficient (ICC) and the ROC curve determining the AUC. Two logistic regression (LR) were used to classify the feature sets, the first to classify

a PZ and the second to detect cancerous regions in the TZ. The CAD had an accuracy of 73% to 86% utilizing the AUC.

Table 4. Features of the studies

Authors	Type of machine learning	Type of diagnosis score	Type of imaging
Fehr <i>et al.</i>	SVM	G-score	T2W
Nguyen <i>et al.</i>	RF	G-score	SLIM
Giannini <i>et al.</i>	SVM	NA	T2W
Peyret <i>et al.</i>	SVM	NA	Multispectral
Wang <i>et al.</i>	SVM	PI-RADS v2	mp-MRI
Cao <i>et al.</i>	CNN	NA	T2W
Reda <i>et al.</i>	SNCAE	NA	DWI
Ginsburg <i>et al.</i>	LR	G-Score	mp-MRI

Some authors [7, 8, 22] use image registration for standardization, given that the multiparametric MRI has to be aligned correctly for this use. Some others instead manipulate the images directly, employing the SLIM imaging technique on prostate biopsy tissue [14]. Moreover, transforming the images from multispectral to grayscale allows a better feature extraction of the image without any standardization for the pixels [15]. The technique based on segment classification by MRI to extract features from each pixel in the image is common to all presented procedures. Machine learning (frequently SVM) will train and later diagnose new images using these features. Then, the system depicts the regions where tumor cells are found [8, 15]. Other approaches implement a Gleason Score to measure malignancy of the prostate cancer [7, 14]. The PI-RADS can also be employed as a grading measure [22].

Table 4 notes the frequent use of the SVM as a machine learning option to develop the topics of interest.

The standard workflow for machine learning implementation is PCa is shown in Fig. 1. [8, 23] begins with the training process. First, the classifier needs the training data that, in this case, are mp-MRI focused on the prostate; the images then are preprocessed, asserting standardization. The next step consists of using feature extraction on the preprocessed images to decide each pixel's most suitable classification method.

When the SVM classifier is trained and a new image is obtained, it has to do the feature extraction and put it into the classifier ending with an estimated segmentation that separates the organ from the surrounding tissue.

Another way to do this process is with a CNN like the case of Cao *et al* [2], which works modifying the data obtained, as seen in Fig. 2, taking the original MR image and its corresponding segmented image in the first instance. The

next step is a convolution that consists of a mathematical operator between a determined filter matrix (in this case is a 3×3 matrix) and the image, getting a new matrix for the feature extraction. Then a max-pooling consisting of a sample window with a specific value (in this case 2×2) runs through the entire image where the value of the higher pixel in the selecting window is recovered and put it into a new matrix of reduced dimensions. The convolution and max-pooling can be repeated many times as need, with the last step being the output layer to give the prostate's estimated segmentation.

4 Discussion

Cancer is the second cause of death globally, only surpassed by cardiovascular diseases. The World Health Organization (WHO) reported 9.6 million casualties worldwide among all cancer types being prostate, the fourth most insidious disease type, accounting for 1.275.106 cases [1]. Prostate cancer (PCa) is the most deadly genre-specific affliction of this type, considering that lung (2.09 million cases), breast (2.09 million cases), and colorectal (1.80 million cases) ranking

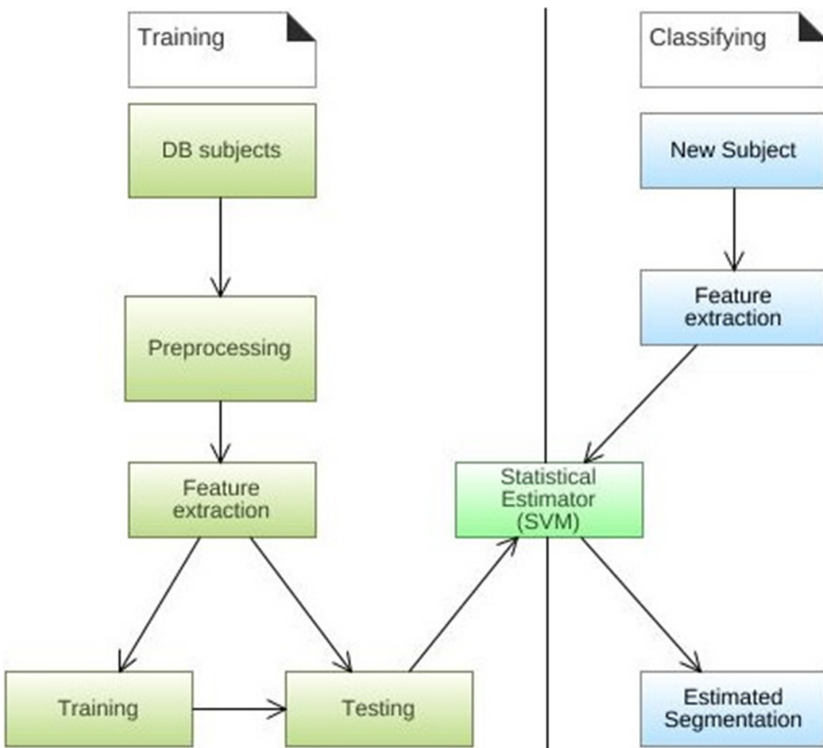


Fig. 1. A Support vector machine model

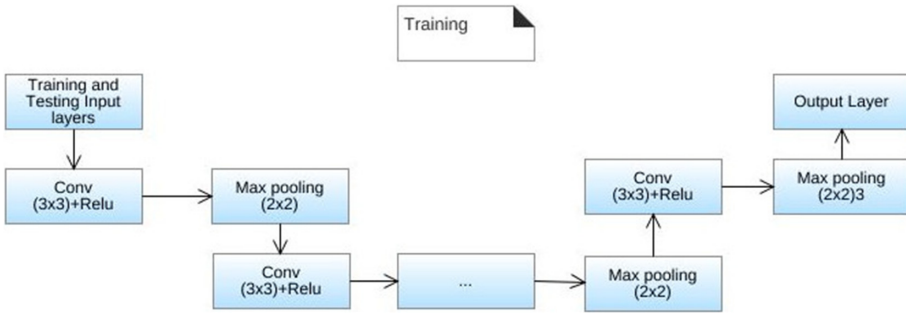


Fig. 2. A convolutional neural network model

in the top three deadly cancers, affect both males and females. Recall that all reported numbers are global for 2018.

Another significant factor in prostate cancer is its independence of socio-economic inputs. A retrospective study showed that men older than 65 established in developed countries have 3.75 times more probability of suffering the disease than men in the same age-range established in non-developed countries [16]. Another remarkable fact is that Afro-descendants are twice more affected than white males, according to global statistics [22].

Although the organ is visible through in-place imaging methods available in developed countries that apport most of the cases in the WHO reports, overdiagnosis has been the reason for 20–40% of the listed cases in Europe and the US [3, 5]. The cited reports point at the Prostate Specific Antigen (PSA) as the cause of the misleading diagnosis. Nevertheless, a significant rate of misdiagnosis is produced by the myriad of highly interpretative factors that specialists should consider, not only while diagnosing but also when classifying the affliction's degree. Exacerbating misdiagnosing causes should include the unusual presentations of PCa [4] that complicate the tasks to experts. The observer's lack of accuracy is not an intuitive concept derived from the inherent human variability; for the specifics, five specialists underwent a blinded test where endorectal mpMRI was provided along with the PIRASDv2 guide that is mastered by these professionals. Although the manuscript concludes a high sensitivity and agreement between operators, the agreement's index reached only 58% for scoring all lesions [10]. Overdiagnosis is a severe concern for healthcare personnel and patients, but it is intrinsic of the curative philosophy implemented in medicine, where indexes are population-based instead of individualized, and the lack of quantifications favors fluctuations in the verdicts. The consequences of overdiagnosis include but are not limited to labeling's psychological and behavioral effects; health detriment secondary to invasive tests, treatment, and follow-up; and financial effects on the overdiagnosed individual and society [19]. Moreover, overtreatment following overdiagnosis can lead to clinically essential consequences, ranging from side effects, e.g., sepsis in a patient undergoing chemotherapy for treatment,

higher rates of myocardial infarction, and suicide have been reported in men with prostate cancer in the year after diagnosis [6,20].

Artificial intelligence (AI) applied to medical imaging has presented a robust alternative to generating clinical verdicts. Machines yield reproducible results, and the capacity to infer rules when we feed them successful experiences as supervisory elements in the learning process makes them capable of overcoming the performance of any automation envisaged before. Authors with expertise in multidisciplinary domains have acknowledged the impact and potential of AI [12,13].

In medical imaging, the AI is mostly used to classify, and thus exert separation of structures in the images, often called segmentation. However, it is ubiquitous to see applications where the automation consists of delivering a verdict. The two approaches mentioned here can appear in the same application, an AI-based segmentation followed by a verdict that uses a supervisory factor, the retrospective diagnosis.

In the case of PCa, the authors have devoted their time to locate lesions using the classifying methods shown in Table 4. The listed classifying methods intend to mimic the PIRADS directives to deliver a grade that could be a Gleason score or any other metric. One can generalize a pipeline where the methods cited in this review can fit to partial or in full extense.

Automatic cancer detection and grading through the image are desired to avoid risky and uncomfortable examination. As mentioned before in this document, artificial intelligence seems to be the right tool to accomplish this complex task. The performed LSR allowed us to know the techniques and current state of art-technology applied to the problem of segmenting the prostate. The authors have used two approaches. The first approximation consists of using a classifier – very often an SVM – fed with features extracted from the image. Then, The classifier is trained to create a separating hyper-plane represented in a statistical estimator. The found hyper-plane exerts separation between the prostate and surrounding tissue. The second approach uses neuronal networks. Here the images are fed to the system, and different features are automatically extracted in a multi-layer implementation. In the two most used approaches, masks manually extracted layer by layer from the training images are used as supervisory elements. Reported accuracy ranges from 88–95%; however, and despite the abundance of implementations intended to segment the prostate automatically, none has reported such instrument being used in the clinics.

To reach a level of implementation, developers should go farther than detecting the prostate's boundaries. An algorithm that detects changes either in the form of the masks or in the organ tissue directly should be in place to yield the numbers needed in further detection and grading pipeline stages. The resulting quantification should be used as features in a new machine learning implementation where the histology results should accomplish the supervision of the learning.

5 Conclusions

Multiple methods are used to determine prostate cancer's presence and its classification as to the degree of malignancy. Unfortunately, various qualitative features are subject to interpretation that in turn can lead to disagreement as to whether treatment is needed, and if so, the scope and method to treat. The use of artificial intelligence with machine learning can provide a more uniform and objective measurement of these tumors using MR imaging characteristics alone, obviating the need for other testing modalities, especially multiple biopsies. Also, the evaluation of treatment outcome can be better assessed with more precise tumor classification. The presented review shows that artificial intelligence is a potent instrument to yield compelling verdicts on prostate cancer diagnosis. Moreover, these machine-produced verdicts are reproducible and render the uncomfortable testing unnecessary.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018)
2. Cao, R., et al.: Joint prostate cancer detection and gleason score prediction in mp-MRI via focalnet. *IEEE Trans. Med. Imaging* **38**(11), 2496–2506 (2019)
3. Draisma, G., et al.: Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J. Natl. Cancer Inst.* **101**(6), 374–383 (2009). <https://doi.org/10.1093/jnci/djp001>
4. Elabbady, A., Kotb, A.F.: Unusual presentations of prostate cancer: a review and case reports. *Arab J. Urol.* **11**(1), 48–53 (2013). <https://doi.org/10.1016/j.aju.2012.10.002>
5. Etzioni, R.: Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *CancerSpectrum Knowl. Environ.* **94**(13), 981–990 (2002). <https://doi.org/10.1093/jnci/94.13.981>
6. Fang, F., et al.: Immediate risk of suicide and cardiovascular death after a prostate cancer diagnosis: cohort study in the United States. *J. Natl Cancer Inst.* **102**(5), 307–314 (2010). <https://doi.org/10.1093/jnci/djp537>
7. Fehr, D.: Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images. *Proc. Natl. Acad. Sci.* **112**(46), E6265–E6273 (2015)
8. Giannini, V.: A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging. *Comput. Med. Imaging Graph.* **46**, 219–226 (2015)
9. Ginsburg, S.B.: Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: preliminary findings from a multi-institutional study. *J. Magn. Reson. Imaging* **46**(1), 184–193 (2017)
10. Greer, M.D., et al.: Accuracy and agreement of PIRADSV2 for prostate cancer mpMRI: a multireader study. *J. Magn. Reson. Imaging* **45**(2), 579–585 (2017)
11. Kitchenham, B.: Procedures for undertaking systematic reviews: Joint technical report. <http://www.inf.ufsc.br/aldo.vw/kitchenham.pdf> (2004)

12. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015). <https://doi.org/10.1038/nature14539>
13. Miller, D.D., Brown, E.W.: Artificial intelligence in medical practice: the question to the answer? *Am. J. Med.* **131**(2), 129–133 (2018). <https://doi.org/10.1016/j.amjmed.2017.10.035>
14. Nguyen, T.H., et al.: Automatic gleason grading of prostate cancer using quantitative phase imaging and machine learning. *J. Biomed. Opt.* **22**(3), 036015 (2017)
15. Peyret, R., Khelifi, F., Bouridane, A., Al-Maadeed, S.: Automatic diagnosis of prostate cancer using multispectral based linear binary pattern bagged codebooks. In: 2017 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART), pp. 1–4. IEEE (2017)
16. Quinn, M., Babb, P.: Patterns and trends in prostate cancer incidence, survival, prevalence and mortality. Part I: international comparisons. *BJU Int.* **90**(2), 162–173 (2002). <https://doi.org/10.1046/j.1464-410X.2002.2822.x>
17. Reda, I., et al.: Computer-aided diagnostic tool for early detection of prostate cancer. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 2668–2672. IEEE (2016)
18. Siegel, R.L., Miller, K.D.: (2018) cancer statistics. *CA Cancer J. Clin.* **68**(1), 7–30 (2018)
19. Singh, H., et al.: Overdiagnosis: causes and consequences in primary health care. *Can. Fam. Physician* **64**(9), 654–659 (2018)
20. Villers, A., Grosclaude, P.: Épidémiologie du cancer de la prostate. Article de revue. *Medecine Nucleaire* **32**(1), 2–4 (2008). <https://doi.org/10.1016/j.mednuc.2007.11.003>
21. Vegega, C., Pytel, P., Pollo-Cattaneo, M.F.: Evaluation of the bias in the management of patient’s appointments in a pediatric office. *ParadigmPlus* **1**(1), 1–21 (2020)
22. Wang, J., Wu, C.J., Bao, M.L., Zhang, J., Wang, X.N., Zhang, Y.D.: 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.* **27**(10), 4082–4090 (2017). <https://doi.org/10.1007/s00330-017-4800-5>
23. Yepes-Calderon, F., Nelson, M.D., McComb, J.G.: Automatically measuring brain ventricular volume within PACS using artificial intelligence. *PLOS ONE* **13**(3), 1–14 (2018). <https://doi.org/10.1371/journal.pone.0193152>