Prostate Cancer Classifier based on Three-Dimensional Magnetic Resonance Imaging and Convolutional Neural Networks

Ana-Maria Minda (Perea) Adriana Albu

Abstract

The main reason for this research is the worldwide existence of a large number of prostate cancers. This article underlines how necessary medical imaging is, in association with artificial intelligence, in early detection of this medical condition. The diagnosis of a patient with prostate cancer is conventionally made based on multiple biopsies, histopathologic tests and other procedures that are time consuming and directly dependent on the experience level of the radiologist. The deep learning algorithms reduce the investigation time and could help medical staff. This work proposes a binary classification algorithm which uses convolutional neural networks to predict whether a 3D MRI scan contains a malignant lesion or not. The provided result can be a starting point in the diagnosis phase. The investigation, however, should be finalized by a human expert.

Keywords: prostate cancer, classification, decision-making, diagnosis, CNN, MRI.

MSC 2020: 68T07, 68T20, 68U10.

1 Introduction

Prostate cancer is the second most diagnosed type of cancer in men, after lung cancer [1], with 1 of 5 cancer cases in most of the countries [2], [3]. The diagnosis is based on biopsy [4], histopathology [1] or measurement of PSA (Prostate Specific Antigen) and DRE (Digital

doi:10.56415/csjm.v31.02

Rectal Examination) in the incipient stages of the disease [3]. Nevertheless, the physicians encounter difficulties because of too many biopsies [4], with low accuracy [5] and specificity [3] in diagnosis. Another problem is the limited medical staff [1].

An MRI (Magnetic Resonance Imaging) scan creates detailed images of organs and tissue based on a magnetic field and radio waves [6]. The use of prostate MRI scans reduces the number of biopsies. Further, convolutional neural networks (CNN) can be trained for automatic processing and interpretation of MRI scans, improving the diagnosis of prostate cancer.

These images contain anatomic and functional parameters (multiparameters MRI, or simple mp-MRI). The use of mp-MRI together with the structured reporting scheme PI-RADS (Prostate Imaging Reporting and Data System) is not quite simple, being strongly dependent on the experience of the radiologist. Therefore, the Computer-aided diagnosis (CAD) systems are a necessity in order to minimize the human effort in the diagnosis procedures [4]. Using deep learning algorithms, the radiologists might obtain an estimation of the severity of this disease, in a shorter time. This would give them the opportunity to focus on treatment and patient support rather than the diagnosis process.

The proposed solution is an algorithm for prostate cancer classification using 3D multi-parameter MRI. Generally, the examination of such an image is made in 30-45 minutes by a human expert. However, using a classification algorithm, the examination is done automatically. The algorithm will classify the prostate lesions as being either benign or malignant.

This work uses PyCharm, Anaconda Development Environment and Jupyter Notebook platform. The algorithm is implemented in Python language, using PyTorch library and MONAI open-source framework for deep learning and healthcare imaging. The classification is made based on DICOM (Digital Imaging and Communications in Medicine) files, which are used to train and test multiple architectures of neural networks. The most performant architecture is included further in the application. The pre-trained neural models were not considered.

2 The context of this research

From a clinical point of view, prostate cancer can be classified as being significant or not significant, according to the aggressivity level of the lesion [2]. The Gleason grade system is generally used to grade prostate cancer lesions. Each investigated tissue receives a Gleason score, which represents the diagnosis of tumor malignancy. This can be low, moderate, or high [1].

Medical images and artificial intelligence mechanisms are frequently used in diagnosis, with undeniable success. This naturally led to the idea of using them in prostate cancer detection [7]. Therefore, two challenges called PROSTATEx have been initiated to speed up the research in this medical field (PROSTATEx Challenge and PROSTATEx-2 Challenge) [8], [9]. A database of 3D MRI scans for prostate cancer has become available, and different algorithms for classification or grading of prostate cancer lesions have been developed. The ground truth set has been defined by physicians with over 20 years of experience, who annotated the images [3].

2.1 Related work

There are multiple studies that explored deep learning algorithms used for prostate cancer detection. For instance, [2], [3], and [7] propose diferent deep learning algorithms that classify prostate lesions. Another study [5] developed an architecture of deep CNN trained on 3D mp-MRI to diagnose prostate cancer. This architecture was inspired by VGG (Visual Geometry Group) networks [5], which use small convolutions (3x3) in order to have a better recognition of the models [2]. The aim of this study was to demonstrate the applicability of deep learning methods in medical imaging for cancer. It was observed that deep learning algorithms perform better than conventional models based on feature engineering [5]. There are also studies that proposed automatic methods for grading the lesions from mp-MRI, using Gleason system (GGG – Gleason Grade Group) [3].

All these studies started based on PROSTATEx challenges (PROSTATEx Challenge and PROSTATEx-2 Challenge) provided by American Association of Physicists in Medicine (AAPM) together with Society of

Photo-Optical Instrumentation Engineers (SPIE) and National Cancer Institute (NCI) [8], [9].

The PROSTATEx-2 challenge aimed to classify prostate cancer using the available image set mp-MRI. One of the solutions used the Ordinal Classification (or Ordinal Regression) technique, which performs a multi-class classification [10]. This solution sorted the lesions according to their aggressivity level $G1 < G2 < G3 < G4 < G5$.

The algorithms developed proved that CNN (VGG-16, for instance) are more efficient when they are used together with other methods, such as Ordinal Classifier. Nevertheless, the solutions have not been tested on other data sets in order to validate their efficiency [2].

Regarding the deep learning algorithms, the easiest way of evaluating them is to analyze the multitude of similar research articles and to run the provided prototypes in the field of interest. Therefore, two public repositories that analyze prostate lesions using the data sets provided by PROSTATEx challenges have been considered.

2.1.1 Lesion classification algorithm using TensorFlow library

Piotr Sobecki proposed a deep learning algorithm for prostate cancer classification based on the MRI data set provided by the first edition of PROSTATEx [11]. The algorithm analyzes the images and classifies the lesions according to their malignity level (the lesions are clinically significant or not). The training data set contains 330 lesions (76 malignant and 254 benign).

The classification algorithm is based on VGG architecture. The repository provides three distinct architectures to analyze the lesions:

- CNN_VGG_SIMPLE – performs a binary classification based on clinical signification of the lesion, using three sub-networks;
- CNN_VGG_MODALITIES – uses a mix of expert architectures, where each modality is able to make predictions;
- CNN_VGG_PIRADS – uses PI-RADS system, together with other expert architectures.

The architectures developed by Piotr Sobecki have the advantage of diversity for each one. On the other hand, there are necessary excessive computations, which lead to a training process that needs too many resources (software, hardware, separated drivers, long time for an epoch, etc.).These architectures have been locally run and some Key Performance Indicators (KPI) have been analyzed. The AUC (Area Under the Curve) values for all three architectures are in the range (0.3, 0.75), with the best results obtained for the architecture CNN_VGG_MODALITIES. These values demonstrate that the algorithm training is slow. In order to improve it, it is necessary to run more than 200 epochs per fold (it is not visible a significant improvement of the KPIs at 100 epochs). Another disadvantage is the old and limited technology (the first version of TensorFlow library was used, which is no longer supported).

2.1.2 Lesion detection and segmentation algorithm using PyTorch framework

This architecture proposes a completely automated system, which takes the mp-MRI prostate scans of a patient suspected of prostate cancer and localizes the lesions using the detection model Retina U-Net. It then performs a segmentation of these lesions and provides the most appropriate Gleason grade (GGG) [12]. This architecture has been developed using the data set provided by the second edition of PROSTATEx challenge. The training set contains 142 lesions (54 of them with GGG 1, 45 with GGG 2, 25 with GGG 3, 10 with GGG 4 and 8 with GGG 5).

The advantages of this architecture are the AUC value of 0.87 [12] and the technologies that were used. On the other hand, it has a limitation because of the Medical Detection Toolkit [13], which is dedicated to Linux operating system only.

2.2 Limitations of the prostate cancer classification

Firstly, the limitation in prostate cancer diagnosis is introduced by MRI scans themselves. The image evaluation is subjectively performed based on PI-RADS criteria or on Gleason system [2]. PI-RADS is a

structured reporting scheme for prostate mp-MRI used to evaluate a patient who is suspected of cancer [14]. But the results are strongly dependent on the human expert who makes the evaluation. This also happens with the Gleason system [2]. Therefore, the deep learning algorithms did not succeed in adequately classify the lesions. In most cases, the tumors were considered low malignant (grade 1) or high malignant (grade 5) [3].

Computer-aided diagnosis (CAD) requires large data sets for training, validation, and testing. The data sets available for prostate cancer are limited because the annotation of MRI scans is tiring and time consuming for human experts. Also, even if multiple patches can be extracted from a single scan, in many cases these are similar, even if the cancer tissues are variate.

Another factor that restricts the use of deep learning in prostate tumors classification is the data imbalance – the observations that belong to individual classes are disproportionate. These observations could not be enough for successfully learning the features of that class [1].

3 Materials and methods

3.1 Available datasets

3.1.1 PROSTATEx Challenge ("SPIE-AAPM-NCI Prostate MR Classification Challenge")

This challenge took place from 21.11.2016 to 15.01.2017 and aimed to determine methods for image analysis in order to diagnose prostate cancer and to classify it from clinical point of view [8], [9], [15], [16]. The data set contains mp-MRI scans of 330 prostate lesions, as well as spatial coordinates, atomic coordinates, and the clinical significance. The testing data set contains 208 prostate lesions.

3.1.2 PROSTATEx-2 Challenge ("SPIE-AAPM-NCI Prostate MR Gleason Grade Group Challenge")

The second challenge took place from 15.05.2017 to 23.06.2017 and focused on some biomarkers for mp-MRI scans, which are useful when

Gleason grade of prostate cancer is determined [8], [9], [15], [16]. Gleason Grade Group (GGG) is a standard for the measurement of the aggressiveness of prostate cancer, which makes possible the prediction of the pathological state and of the oncological result. A GGG grade is assigned to a possible lesion using histopathological analysis of biopsies. The GGG grades are natural numbers between 1 and 5, where 1 is a lesion that doesn't require treatment and 5 is the most severe prostate cancer lesion [1], [3].

The training set contains 112 prostate lesions, the spatial and atomic coordinates, as well as the GGG grade of each lesion. The testing set contains 70 lesions.

3.1.3 DiagSet

Besides the data sets provided by PROSTATEx challenges, there are also other data sets available that can be used to train deep learning algorithms. DiagSet is a histopathologic data set, which contains regions of prostate tissue, annotated in WSI (Whole Slide Image) scans. These regions have been marked with GGG grade, in order to specify the severity level of the disease. The annotation was made by human experts in histopathology [1].

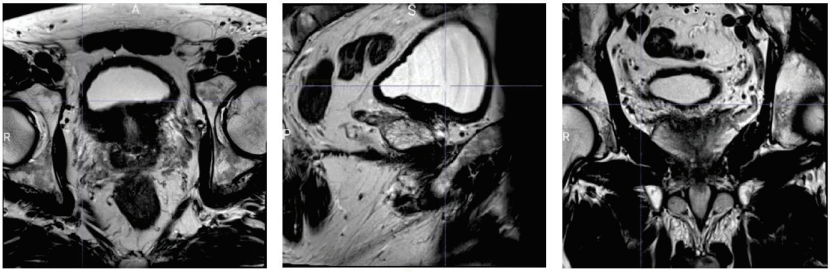
This data set was used to train several CNN architectures such as AlexNet, VGG16, VGG19, ResNet50, and InceptionV3 in different configurations or different dimensions of the images. The best configurations have been used to create an ensemble of classifiers. The obtained results have been validated by physicians [1].

3.2 Image types

The solution presented in this work used the data set provided by PROSTATEx challenge ("SPIE-AAPM-NCI Prostate MR Classification Challenge") [8], [9], [15], [16]. The main reason for this is that at least five 3D MRI sequence types are provided for each patient. The ground truth is available, and it is used in both the training and evaluation processes.

There are three anatomic planes that compound a 3D MRI: transverse plane (is perpendicular to the spine and divides the body into

superior and inferior parts), sagittal plane (divides the body into left and right parts), and coronal plane (divides the body in vertical plane into dorsal and ventral parts) [17]. Fig. 1 [8], [9] presents the scans for the three anatomic planes of a patient from training data set PROSTATEx.



a b c

Figure 1. (a) transverse plane, (b) sagittal plane, and (c) coronal plane

The five sequence types [7] that are present in 3D MRI of all patients are:

- Transverse sequences T2W (T2-wieghted);
- Sagittal sequences T2W;
- Transverse sequences ADC (Apparent Diffusion Coefficient);
- Transverse sequences DWI (Diffusion-Weighted Imaging);
- Transverse sequences Ktrans (a measure for capillary permeability).

The four types of transverse images (T2W, ADC, DWI, and Ktrans) are presented in Fig. 2 [8], [9] (they were also selected from training data set PROSTATEx).

All sequences have been compressed in DICOM files, and then provided as a public data set in PROSTATEx challenge. DICOM (Digital Imaging and Communications in Medicine) is an international standard that defines medical image formats, being one of the most frequently

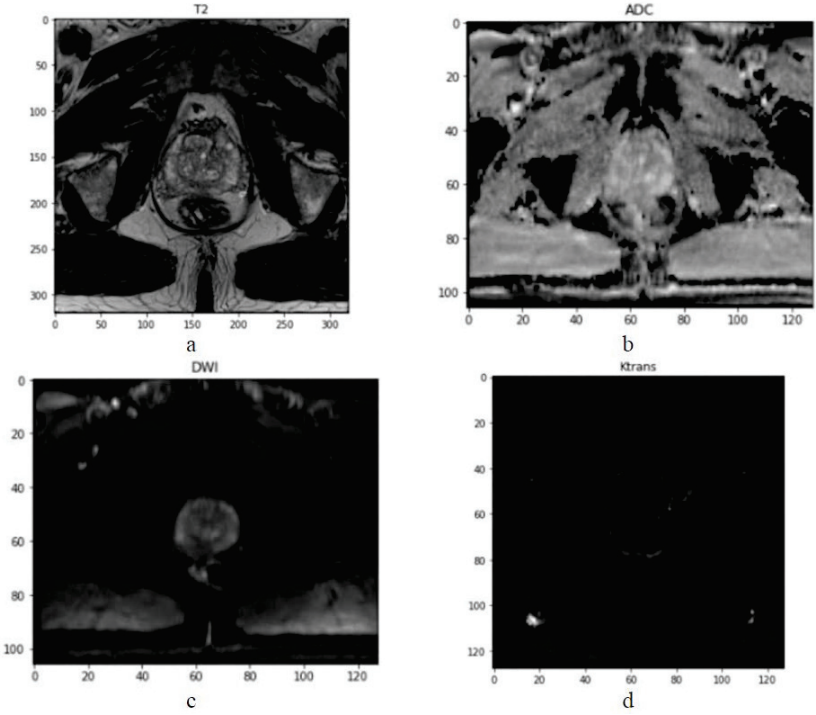


Figure 2. (a) a T2W slice, (b) an ADC slice, (c) a DWI slice, and (d) a Ktrans slice

used standards for medical information transfer. DICOM is implemented in almost all equipment from radiology, radiotherapy, cardiologic imaging, and other medical domains such as ophthalmology or stomatology [18]. DICOM standard is recognized by the International Organization for Standardization as standard ISO 12052 [18].

The current project also uses NIfTI (Neuroimaging Informatics Technology Initiative) files. This format is dedicated to neural imaging, having two versions: NIfTI-1 and NIfTi-2 (which is an update of NIfTI-1 that allows storing a larger quantity of data) [19].

3.3 Development environments

The following technologies were used throughout the development process:

- Anaconda.org - service for packages management;
- Pip and Conda - systems for packages management;
- PyCharm - development environment;
- Jupyter Notebook - application used for a visual perspective of the data and the proposed solution.

The programming language used for the entire project is Python, in addition to a couple of frameworks for training, validation, and testing the model (PyTorch and MONAI) and several libraries for data reading and visualization (Pandas, Matplotlib, Nibabel).

3.4 System's architecture

The proposed solution has a simple architecture (Fig. 3). The MRI scans (DICOM format) of the data set provided by PROSTATEx challenge are locally processed. The operating system of the workstation is Windows. Once the images are transformed into NIfTI format, they are processed using a DenseNet-121 classification algorithm. The output data of the training phase are the parameters of the model (the best results of the training). Then, the model can be evaluated (testing phase).

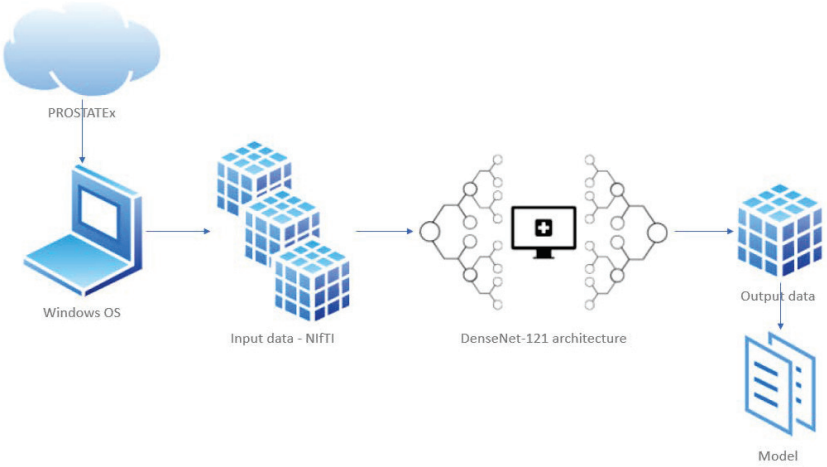


Figure 3. The architecture of the proposed solution

The main steps that were followed in the development of this solution are:

- Analysis of the existing repositories;
- Virtual environment setup for the Python packages and libraries;
- Available data (images and labels) visualization and analysis;
- Data processing;
- Training of the classification algorithm;
- Evaluation of the classification algorithm.

The classification algorithm uses a DenseNet-121 network, with the following features: one convolution 7x7, 58 convolutions 3x3, 61 convolutions 1x1, 4 average pooling layers and one fully connected layer [20]. In a DenseNet network, each layer is directly connected to the next layer (Densely Connected Convolutional Network). Therefore, for n layers, there will be $n(n+1)/2$ direct connections [20]. According to the MONAI documentation, because the data set images are three-dimensional, the DenseNet-121 is not pre-trained.

3.5 Model training and evaluation

The model was trained, validated, and tested using the PyTorch framework. In order to perform this procedure (specific to deep learning algorithms), the data set provided by PROSTATEx challenge has been randomly divided into sub-sets for training, validation, and testing. The data set contains 204 patients. Training and validation phases are performed using 80% of the patients (80% of them for training and 20% for validation). The rest of the patients (20% of the initial set) remain for testing phase. Therefore, 163 patients are used for training and validation (80% of 204) and 41 for testing (20% of 204).

Each image of these sub-sets has a label that specifies whether a lesion is malignant (clinically significant) or not. The labels were set by experienced radiologists.

Once the dataset is split and labeled, Transforms objects are defined. These objects scale the intensity of the images, re-dimension them and rotate them by 90°. The data type will also be verified, in order to ensure that these are PyTorch tensors [21]. Each sub-set has a DataLoader which contains an object of type ImageDataset and hardware specifications to parallelize the training, validation, and testing processes [22]. The ImageDataset objects contain a single-dimensional matrix for the images of the sub-set, a 1D matrix for the labels and a Transforms object.

After the DataLoader objects are created, the model can be trained, validated, and tested. Several parameters such as the number of parallel processed images (batch_size), the number of parallel threads (num_workers), or number of epochs for training phase can be set in a Python configuration file.

The training phase uses a simple PyTorch loop, which evaluates in each iteration the internal performance of the model using the validation sub-set. The epochs' parameters that improve the metrics are saved, being used in the final model, which is the best model produced.

At the end, the model is evaluated comparing the predictions with the labels. This will determine key performance indicators (KPI) based on the elements of the confusion matrix:

- TP (true positive) – both algorithm and label indicate a malig-

nant lesion;

- TN (true negative) – both algorithm and label indicate a benign lesion;

- FP (false positive) – the algorithm predicts a malignant lesion, but the label indicates a benign lesion;

- FN (false negative) – the algorithm predicts a benign lesion, but the label indicates a malignant lesion.

The KPIs that are considered in model evaluation are [23]:

- Accuracy – the most frequently used metric when an algorithm is evaluated, indicating the number of correct predictions reported to the total predictions (Eq. (1)):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}; \tag{1}$$

- Precision – the report between correct positive predictions and all positive predictions (Eq. (2)):

$$Precision = \frac{TP}{TP + FP}; \tag{2}$$

- Recall (sensibility) – measures how many positive cases have been correctly predicted reported to all positive cases of the data set (Eq. (3)):

$$Recall = \frac{TP}{TP + FN}; \tag{3}$$

- F1-Score – the harmonic average of precision and recall (Eq. (4)):

$$F1\text{-}Score = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{4}$$

4 Experimental results and discussions

Fig. 4 represents a DataFrame object which contains information and results from the training and testing phases of the classification algorithm. The "color", "loss_function", "optimizer", "learning_rate", and "epochs" columns define the color code of the graphical representation, the loss function type, the optimizer type, and the learning rate, as well as the number of the epochs assigned to the classification algorithm. The "best_accuracy" column represents the best value obtained in the training phase. The last 8 columns are the KPI values obtained in the testing phase.

i	color	loss_function	optimizer	learning_rate	epochs	best_accuracy	TP	TN	FP	FN	accuracy	recall	precision	f1
0	red	Binary Cross Entropy	Adam	1e-3	100	0.7214	0	174	0	81	0.682	0.000	-0.100	-0.100
1	lightgreen	Binary Cross Entropy	Adam	1e-4	50	0.6990	0	136	0	126	0.519	0.000	-0.100	-0.100
2	blue	Binary Cross Entropy	Adam	1e-4	50	0.6768	26	160	12	64	0.710	0.289	0.684	0.406
3	yellow	Binary Cross Entropy	Adam	1e-4	100	0.6318	22	153	22	59	0.684	0.272	0.500	0.352
4	lightgray	Binary Cross Entropy	Adam	1e-4	200	0.7114	6	164	1	80	0.677	0.070	0.857	0.129
5	pink	Binary Cross Entropy	Adam	1e-4	50	0.5149	2	177	4	86	0.665	0.023	0.333	0.043
6	orange	Binary Cross Entropy	Adam	1e-4	50	0.8146	5	146	4	105	0.581	0.045	0.556	0.084
7	brown	Binary Cross Entropy	Adam	1e-5	50	0.6745	21	126	25	82	0.579	0.204	0.457	0.282
8	purple	Binary Cross Entropy	Adam	1e-5	200	0.7608	9	154	8	84	0.639	0.097	0.529	0.164

Figure 4. Information obtained during the training and testing process

Figs. 5, 6, and 7 present a visual representation of the internal results during the training process. The values of the loss function are directly related to the type of the function, to the number of epochs, to the optimizer type, and to the learning rate value. For instance, the larger the number of epochs, the smaller the value of the loss function – in Fig. 5, the loss function is around 0.2, while in Figs. 6 and 7, the values are less than 0.1.

Another aspect that can be observed in Figs. 5, 6, and 7 is that the internal accuracy of the algorithm in the validation stage reaches the highest values in the first 50 training epochs. Therefore, the classification algorithm doesn't need a larger number of iterations to determine the optimal values of the parameters.

The best values of the evaluated models are represented in Fig. 8. During the training process, the model with the best results is the orange one (i=6, with an accuracy of over 80%) with the following fea-

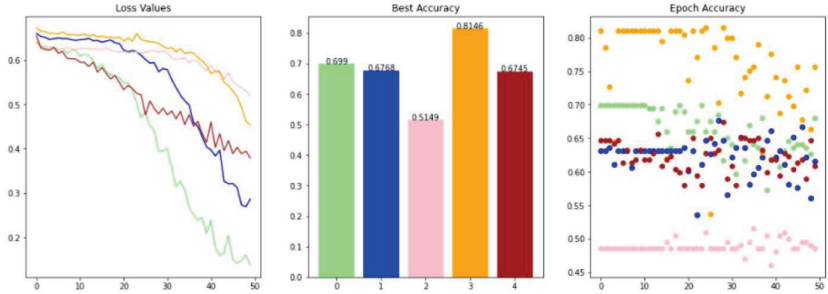


Figure 5. Loss function values (left), accuracy values (right), and the best accuracies of the models (middle) in the training phase – 50 epochs

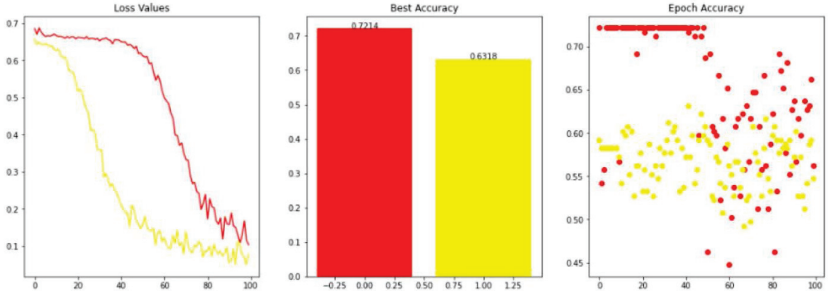


Figure 6. Loss function values (left), accuracy values (right), and the best accuracies of the models (middle) in the training phase – 100 epochs

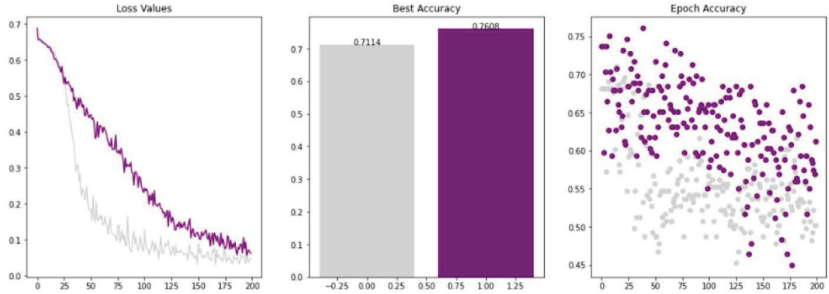


Figure 7. Loss function values (left), accuracy values (right), and the best accuracies of the models (middle) in the training phase – 200 epochs

tures: 50 epochs, Binary Cross Entropy loss function, Adam optimizer, and 1e-4 learning rate. Figs. 9 and 10 represent the KPI values ob-

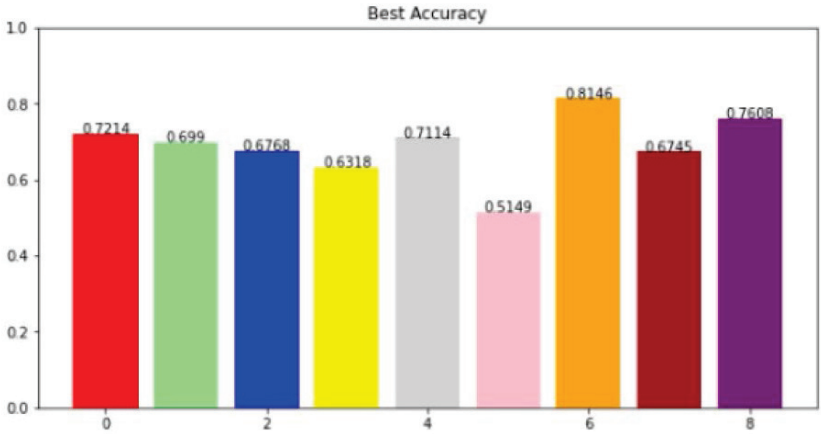


Figure 8. The best accuracy results in the training phase

tained in the testing phase. According to these results, the models that successfully identified malignant lesions are the blue one (i=2) and the yellow one (i=3). Nevertheless, the yellow model is in top of the models with the most FP values. Therefore, the blue model can be considered the most suitable for prostate lesions classification, according to the KPI values.

Looking at the representation of the metrics, the best accuracy is provided by the blue model (i=2), with over 70% correct predictions. Other models that correctly classified over 65% of the lesions are red (i=0), yellow (i=3), light grey (i=4), and pink (i=5). Regarding the recall and F1-score, the models blue and yellow have the best results. The precision has the highest values for the models light grey and blue. Nevertheless, the recall and F1-score values of the light grey mode are much smaller than those of the blue model. Therefore, the blue model (i=2) is the best model once the testing phase of the classification algorithm is completed.

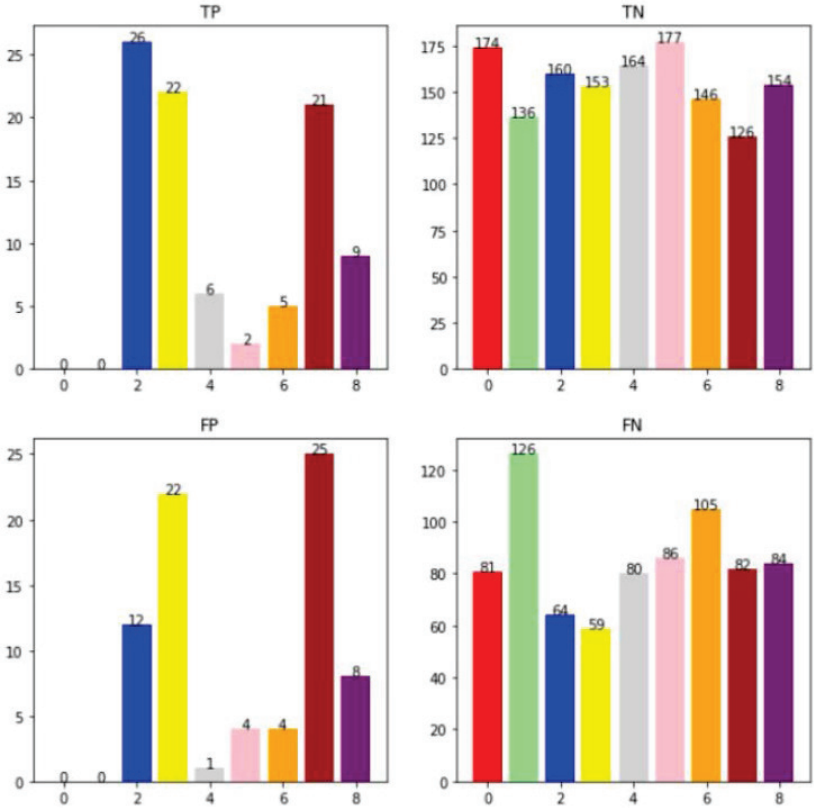


Figure 9. The elements of the confusion matrix in the testing phase

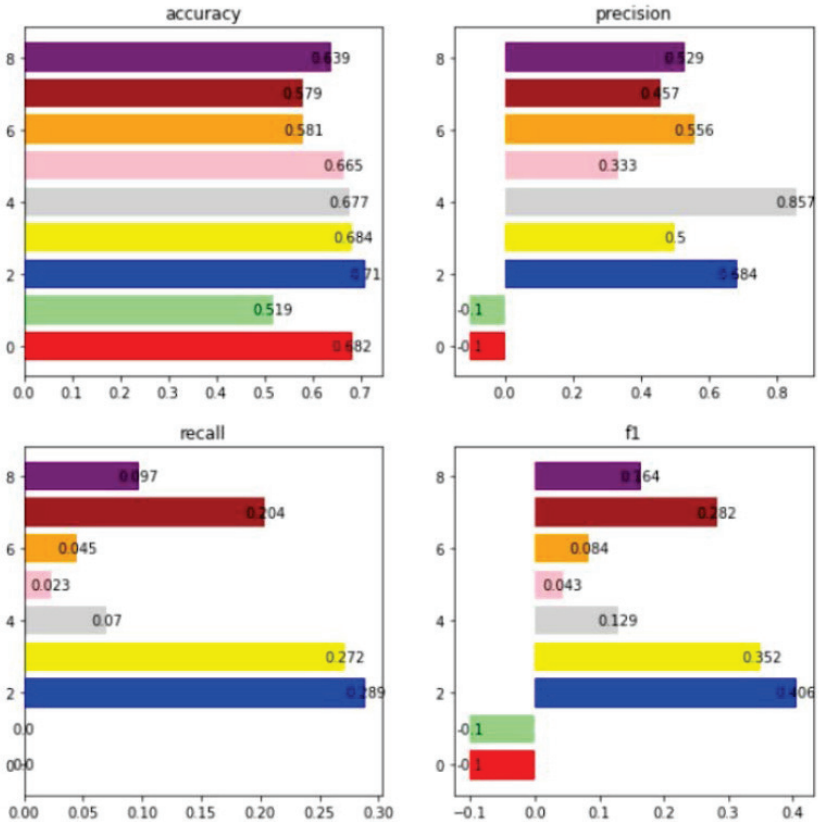


Figure 10. The performance metrics in the testing phase

4.1 Special aspects observed

During the evaluation of the classification algorithm, different values and trends of the KPI and of the loss function have been compared. The aim of the analysis of these values is to set the correct values and to determine the best parameters of the model.

The oscillations in the loss function indicate that learning rate is either too small or too large, even if it decreases exponentially during the epochs. The solution for such a problem is to train the classification algorithm with different optimizers, architectures, loss functions,

and learning rates. The process of finding the best parameters is an experimental one. Through the metrics analysis in the testing phase, the following can be observed:

- Precision – missing positive predictions (TP+FP=0). The precision in the red and light green models resulted in negative values. These negative values have been assigned to avoid and to mark a division by zero (during the testing process, no lesion was classified as malignant).
- Recall – missing correct positive predictions (TP=0). The recall was zero in the red and light green models. This means that no malignant lesion was correctly classified during the testing process.
- F1-score – irrelevant. This is directly dependent on precision and recall; as a result, for the previous two cases, the F1-score becomes irrelevant.

Therefore, all these three metrics (precision, recall and F1-score) are directly dependent on the number of positive predictions in the testing phase. The number of malignant lesions of the prostate in the data set is small. The solution in this direction is to increase the number of malignant lesions in the training sub-set by an oversampling process (a technique for the adjustment of a class distribution in a data set [24]).

5 Conclusion

This work proposes a binary classification algorithm for prostate lesions. It predicts whether an MRI scan contains a malignant lesion or not. The images and labels provided by PROSTATEx challenge have been analyzed. A new architecture, which uses current technologies (such as PyTorch and MONAI frameworks) has been created. Several models have been evaluated and the most performant architecture has been selected to present the performances of the classification algorithm.

The advantages of the proposed solution are given by the current technologies (which have constant technical support) and by the optimizations in the training and evaluation phases. The PyTorch frame-

work does not require the installation of additional drivers in order to use GPU. The analysis and the processing of the MRI sequences is faster using GPU. Unlike TensorFlow, this framework does not overcharge the hardware resources. Thus, models can be obtained with minimum degradation of the physical resources that are used.

Another advantage is the compatibility of the MONAI framework with multiple operating systems. The solution is easily configurable and accessible to the developers that lead research projects in this area. The source code is available on GitHub [25].

The solution provided by this research is the first step in development of a deep learning application for prostate lesions classification. The next steps involve the integration of this model into an application that provides a graphical user interface. The physicians might upload a 3D MRI into the application, obtaining a classification into "malignant lesion" or "benign lesion". More than this, classification algorithms are the simplest method of analyzing 3D medical images. The development of algorithms for detection and segmentation of the lesions would improve the application.

References

[1] M. Koziarski, B. Cyganek, B. Olborski, Z. Antosz, M. Zydak, B. Kwolek, P. Wasowicz, A. Bukała, J. Swadzb, and P. Sitkowski, "DiagSet: a dataset for prostate cancer histopathological image classification," *arXiv, Electrical Engineering and Systems Science, Image and Video Processing*, May 2021. [Online]. Available: https://arxiv.org/pdf/2105.04014.pdf.

[2] B. Abraham and M.S. Nair, "Automated Grading of Prostate Cancer using Convolutional Neural Network and Ordinal Class Classifier," *Informatics in Medicine Unlocked*, Elsevier, vol. 17, no. 1, Article no. 100256, October 2019.

[3] C. de Vente, P. Vos, M. Hosseinzadeh, J. Pluim, and M. Veta, "Deep Learning Regression for Prostate Cancer Detection and Grading in Bi-Parametric MRI," *IEEE Trans. Biomedical Engineering*, vol. 68, no. 2, pp. 374–383, Feb. 2021.

[4] S.G. Armato, H. Huisman, K. Drukker, L. Hadjiiski, J.S. Kirby, N. Petrick, G. Redmond, M.L. Giger, K. Cha, A. Mamonov, J. Kalpathy-Cramer, and K. Farahanif, "PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images," *Journal of Medical Imaging*, vol. 5, no. 4, Article no. 044501, Bellingham, 2018.

[5] S. Liu, H. Zheng, Y. Fengc, and W. Li, "Prostate Cancer Diagnosis using Deep Learning with 3D Multiparametric MRI," in *Proceedings of SPIE Medical Imaging, vol. 10134*, (Orlando, Florida, United States), 2017. [Online]. Available: https://doi.org/10.1117/12.2277121.

[6] P. Lam and J. Marcin, "What to know about MRI scans," *Medical News Today*, 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/146309. Accessed on: Jan. 2022.

[7] Hugegene, "3D CNN Classification of Prostate Cancer on PROSTATEx-2," *Towards Data Science*, 2019. [Online]. Available: https://towardsdatascience.com/3d-cnn-classification-of-prostate-tumour-on-multi-parametric-mri-sequences-prostatex-2-cced525394bb.

[8] "ProstateX challenge," *Grand Challenge*. [Online]. Available: https://prostatex.grand-challenge.org/. Accessed on: January 2022.

[9] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "ProstateX Challenge data," *The Cancer Imaging Archive*, 2017. [Online]. Available: https://doi.org/10.7937/K9TCIA.2017.MURS5CL. Accessed on: Nov. 2021.

[10] L. Gaudette and N. Japkowicz, "Evaluation Methods for Ordinal Classification," *Advances in Artificial Intelligence, Canadian AI 2009* (Lecture Notes in Computer Science, vol. 5549), Berlin, Heidelberg: Springer, 2009, pp 207–210.

[11] P. Sobecki, "Prostate lesion classification using Deep Convolutional Neural Networks," GitHub. [Online]. Available: https://github.com/piotrsobecki/PCa-CNNs, 2020. Accessed on: Nov. 2021.

[12] O.J. Pellicer-Valero, J.L. Marenco Jiménez, V. Gonzalez-Perez, J.L.C. Ramón-Borja, I.M. García, M. Barrios Benito, P. Pelechano Gómez, J. Rubio-Briones, M. José Rupérez, and J.D. Martín-Guerrero, "Deep Learning for fully automatic detection, segmentation, and Gleason Grade estimation of prostate cancer in multiparametric Magnetic Resonance Images," *Scientific Reports*, vol. 12, Article no. 2975, 2022.

[13] P.F. Jaeger, S.A.A. Kohl, S. Bickelhaupt, F. Isensee, T. Anselm Kuder, H.-P. Schlemmer, and K.H. Maier-Hein, "Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection," *arXiv, Computer Science, Computer Vision and Pattern Recognition*, Nov. 2018. [Online]. Available: https://arxiv.org/pdf/1811.08661.pdf.

[14] M. Czarniecki and Y. Weerakkody, "Prostate Imaging-Reporting and Data System (PI-RADS)," *Radiopaedia.org*, 2021. [Online]. Available: https://doi.org/10.53347/rID-27968. Accessed on: Jan. 2022.

[15] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in MRI," *IEEE Trans. Medical Imaging*, vol. 33, pp.1083–1092, 2014. Available: https://doi.org/10.1109/TMI.2014.2303821. Accessed on: Nov. 2021.

[16] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior, "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository," *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013, Available: https://doi.org/10.1007/s10278-013-9622-7. Accessed on: Nov. 2021.

[17] "1.4D: Body Planes and Sections," in *Anatomy and Physiology (Boundless), 1: Introduction to Anatomy and Physiology*, LibreTexts, Medicine, 2020.

[18] "About DICOM: Overview," *DICOM – Digital Imaging and Communications in Medicine*. [Online]. Available: https://www.dicomstandard.org/about-home. Accessed on: Jul. 2022.

[19] C. Moore and H. Knipe, "NIfTI (file format)," *Radiopaedia.org*, 2014. [Online]. Available: https://doi.org/10.53347/rID-72562. Accessed on: Jul. 2022.

[20] A. Ahmed, "Architecture of DenseNet-121," *OpenGenus IQ: Computing Expertise & Legacy, Machine Learning (ML) Tutorial*. [Online]. Available: https://iq.opengenus.org/architecture-of-densenet121/. Accessed on: Jul. 2022.

[21] "Transforms," *MONAI*. [Online]. Available: https://docs.monai.io/en/stable/transforms.html. Accessed on: Jul. 2022.

[22] "Datasets & DataLoaders," *PyTorch*. [Online]. Available: https://pytorch.org/tutorials/beginner/basics/data_tutorial.html. Accessed on: Jul. 2022.

[23] T. Kanstrén, "A Look at Precision, Recall, and F1-Score," *Towards Data Science*, 2020. [Online]. Available: https://towardsdatascience.com/a-look-at-precision-recall-and-f1-score-36b5fd0dd3ec. Accessed on: Jul. 2022.

[24] N.V. Chawla, "Data Mining for Imbalanced Datasets: An Overview," in *Data Mining and Knowledge Discovery Handbook*, Springer, 2010, pp. 875–886. ISBN 978-0-387-09823-4. Available: https://doi.org/10.1007/978-0-387-09823-4_45.

[25] A.-M. Minda (Perea), "Prostate Cancer Classifier," GitHub, Jul. 2022. [Online]. Available: https://github.com/pereaanamaria/prostate_cancer_classifier.

Ana-Maria Minda (Perea), Adriana Albu

Received November 01, 2022
Revised February 20, 2023
Accepted February 21, 2023

Ana-Maria Minda (Perea)
ORCID: https://orcid.org/0009-0008-1415-4509
Faculty of Automation and Computers,
Politehnica University Timisoara
Timisoara, Romania
E–mail: anamariaperea10@gmail.com

Adriana ALBU
ORCID: https://orcid.org/0000-0003-1579-6163
Department of Automation and Applied Informatics,
Politehnica University Timisoara
Timisoara, Romania
E–mail: adriana.albu@upt.ro