applied sciences

MDPI

Article

A Hybrid End-to-End Approach Integrating Conditional Random Fields into CNNs for Prostate Cancer Detection on MRI

Paulo Lapa [1], Mauro Castelli [1,*], Ivo Gonçalves [2], Evis Sala [3,4] and Leonardo Rundo [3,4]

1 Nova Information Management School (NOVA IMS), Campus de Campolide, Universidade Nova de Lisboa, 1070-332 Lisboa, Portugal; plapa@novaims.unl.pt
2 INESC Coimbra, DEEC, University of Coimbra, Pólo 2, 3030-290 Coimbra, Portugal; ivogoncalves77@gmail.com
3 Department of Radiology, University of Cambridge, Cambridge CB2 0QQ, UK; es220@cam.ac.uk (E.S.); lr495@cam.ac.uk (L.R.)
4 Cancer Research UK Cambridge Centre, Cambridge CB2 0RE, UK
* Correspondence: mcastelli@novaims.unl.pt; Tel.: +351-2138-2861-0208

Received: 6 October 2019; Accepted: 24 December 2019; Published: 2 January 2020

Featured Application: Integration of Conditional Random Fields into Convolutional Neural Networks as a hybrid end-to-end approach for prostate cancer detection on non-contrast-enhanced Magnetic Resonance Imaging.

Abstract: Prostate Cancer (PCa) is the most common oncological disease in Western men. Even though a growing effort has been carried out by the scientific community in recent years, accurate and reliable automated PCa detection methods on multiparametric Magnetic Resonance Imaging (mpMRI) are still a compelling issue. In this work, a Deep Neural Network architecture is developed for the task of classifying clinically significant PCa on non-contrast-enhanced MR images. In particular, we propose the use of Conditional Random Fields as a Recurrent Neural Network (CRF-RNN) to enhance the classification performance of XmasNet, a Convolutional Neural Network (CNN) architecture specifically tailored to the PROSTATEx17 Challenge. The devised approach builds a hybrid end-to-end trainable network, CRF-XmasNet, composed of an initial CNN component performing feature extraction and a CRF-based probabilistic graphical model component for structured prediction, without the need for two separate training procedures. Experimental results show the suitability of this method in terms of classification accuracy and training time, even though the high-variability of the observed results must be reduced before transferring the resulting architecture to a clinical environment. Interestingly, the use of CRFs as a separate postprocessing method achieves significantly lower performance with respect to the proposed hybrid end-to-end approach. The proposed hybrid end-to-end CRF-RNN approach yields excellent peak performance for all the CNN architectures taken into account, but it shows a high-variability, thus requiring future investigation on the integration of CRFs into a CNN.

Keywords: prostate cancer detection; magnetic resonance imaging; convolutional neural networks; conditional random fields; recurrent neural networks

1. Introduction

According to the American Cancer Society, Prostate Cancer (PCa) is the most common type of cancer in Western men [1]; in 2018, approximately 1.3 million new cases were diagnosed and 359,000 related deaths occurred worldwide [2]. Despite its incidence and societal impact, the current diagnostic

techniques—i.e., Digital Rectal Exam, Prostate-Specific Antigen (PSA) [3]—may be subjective and error-prone [4]. Furthermore, intra-tumor heterogeneity is observed in PCa, contributing to disease progression [5].

Currently, high-resolution multiparametric Magnetic Resonance Imaging (mpMRI) for Computer-Aided Diagnosis (CAD) is gaining clinical and scientific interest [6] by enabling quantitative measurements for intra- and inter-tumoral heterogeneity based on radiomics studies [7]. Additional and often complementary information can be acquired by means of different MRI sequences: anatomical information can be obtained using T2-weighted (T2w), T1-weighted (T1w) and Proton Density (PDw) protocols [4,8,9]. Further information is conveyed by functional imaging [10], allowing for better depiction of multiple aspects of the tumor structure: its micro-environment by estimating the water molecule movement using Diffusion-Weighted Imaging (DWI) and the derived Apparent Diffusion Coefficient (ADC) maps [11], as well as the vascular structure of the tumor with Dynamic Contrast-Enhanced (DCE) MRI [12]. Unfortunately, multi-focal tumors in the prostate occur commonly, posing additional challenges for accurate prognoses on MRI [13]; thus, devising and exploiting advanced Machine Learning methods [14,15] for prostate cancer detection and differentiation is clinically relevant [16].

Therefore, the tasks of PCa classification can benefit from the combination of several modalities, each conveying clinically useful information [8,17]. Clinical consensus for PCa diagnosis typically considers mpMRI by combining T2w with at least two functional imaging protocols [18]. In this work, T2w, PDw and ADC MRI sequences were chosen as inputs for the models. T2w conveys relevant information about the prostate zonal anatomy [15] as well as tumor location and extent [19]: PCa has low signal intensity, which can be suitably detected from the healthy hyper-intense peripheral zone tissue (harboring approximately 70% of PCa cases [20]), but it is more difficult to differentiate in the central and transitional zones due to their normal low signal intensity [4,8].

PDw, by quantifying the amount of water protons contributing to each voxel [9], provides a good distinction between fat and fluid [21]. ADC yields a quantitative map of the water diffusion characteristics of the prostatic tissue: PCa typically has packed and dense regions with intra- and inter-cellular membranes that influence water motion [4,8]. Lastly, DCE sequences depict the patient's vascular system in detail, since tumors exhibit a highly vascularized micro-environment, by exploiting a Gadolinium-based contrast medium [22].

In Deep Learning applications to medical image analysis, some challenges are still present [23]; namely, (i) the lack of large training data sets, (ii) the absence of reliable ground truth data, and (*iii*) the difficulty in training large models [24]. Nonetheless, some factors are consistently present in successful models [25]: expert knowledge, novel data preprocessing or augmentation techniques, and the application of task-specific architectures.

Despite the growing interest in developing novel models for the task of PCa, little effort has been devoted to the addition of new types of layers. Recently, the Semantic Learning Machine (SLM) [26–28] neuroevolution algorithm was successfully employed to replace the backpropagation algorithm commonly used in the Fully-Connected (FC) layers of Convolutional Neural Networks (CNNs) [29,30]. When compared with backpropagation, SLM achieved higher classification accuracy in PCa detection as well as a training speed-up of one order of magnitude. A CNN architecture developed for the classification task is typically composed of regular layers, which perform convolutions, dot products, batch normalization, or pooling operations. This work integrates a Conditional Random Field (CRF) model as a Recurrent Neural Network (RNN) [31], generally exploited in segmentation, to the PCa classification problem. In accord with the latest clinical trends aiming at decreasing contrast medium usage [4], we analyzed only the non-contrast-enhanced mpMRI sequences to assess also the feasibility of our methodology from a patient safety and health economics perspective [32].

Research Questions.

We specifically address two questions:

- Can the CRF-CNN be integrated into a state-of-the-art CNN as an end-to-end approach?
- Can the smoothing effect of CRFs increase the classification performance of CNNs in PCa detection?

Contributions.

Our main contributions are the following:

- A hybrid end-to-end trainable network that combines CRF-RNN [31] and a state-of-the-art CNN, namely XmasNet [33], without requiring a two-phase training procedure.
- The proposed CRF-XmasNet architecture generally outperforms the baseline architecture XmasNet [33] in terms of PCa classification on mpMRI.
- The proposed end-to-end integration of CRF-RNN provides better generalization ability when compared to a two-phase implementation, using a CRF as a postprocessing step.

In particular, the proposed approach aimed at outperforming XmasNet, a network specifically created for dealing with prostate cancer MRI data. This network is the most state-of-the-art for this kind of application. Thus, outperforming it would be a valuable contribution in the medical field, as it shows how the integration of CRFs in XmasNet could improve the performance of the commonly used XmasNet architecture.

The manuscript is organized as follows. Section 2 introduces the theoretical foundations of CRFs and the Deep Neural Network (DNN) architectures underlying the devised method. Section 3 presents the characteristics of the analyzed prostate mpMRI dataset, as well as the proposed method. Section 4 shows and critically analyzes the achieved experimental results. Finally, Section 5 concludes the paper and suggests future research avenues.

2. Theoretical Background

This section introduces the basic concepts necessary to fully understand the rationale and the functioning of the devised DNN for PCa detection.

2.1. Convolutional Neural Networks

CNNs have become one of the most common supervised learning techniques [34]. They can learn complex patterns from unstructured data (i.e., text or images) with limited domain knowledge. By leveraging the convolution operation, CNNs can perform their task on two- or higher-dimensional inputs. They can consider the neighboring region around a pixel, making them well-suited for image applications [35].

They have found success in Medical Image Analysis (MIA) applications, namely in cancer-related problems [25]. In the prostate region, deep CNNs achieved better performance when compared with non-deep CNNs [36] for PCa classification tasks. For this task, CNNs have also been used to extract discriminative features from T1w and DCE sequences [37], from 3D features extracted either from MRI sequences [38] or Gleason Score (GS) prediction based on Transrectal Ultrasound (TRUS)-guided biopsy results [39,40]. CNNs have also been used with U-Net inspired architectures [41,42] for PCa segmentation.

At the most general level, CNNs have been used in almost every anatomic area of the human body (e.g., brain, eyes, torso, knees) for various tasks (disease location, tissue segmentation or survival probability calculation) with a high degree of success [25].

For instance, XmasNet was developed by Liu et al. [33] specifically for the PROSTATEx Challenge 2017 [43], inspired by the Visual Geometry Group (VGG) net [44]. Despite its relative simplicity, it achieved state-of-the-art results, outperforming 69 methods of 33 groups and having the second highest Area Under the Receiver Operating Characteristics Curve (AUROC) on the unseen test

set. XmasNet is a relatively traditional architecture: four convolutional layers and two FC layers. The architecture also makes use of Batch Normalization, Rectified Linear Unit (ReLU) activation functions, and Max Pooling.

Other architectures were compared in this work, namely AlexNet [34], VGG16 [44] and ResNet [45]. AlexNet was the winner of the ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [46], and its success revived the interest in CNNs in Computer Vision and brought about several novel implementations: ReLU non-linear activation functions, multiple GPU training, Local Response Normalization, and Overlapping Pooling, that are still used in current architectures. VGG16 is based on AlexNet and was the first truly deep CNN with 16 convolutional layers, made possible by the use of small convolutional filters and ReLU activation functions whenever possible [34]. It achieved first and second place in ILSVRC 2014 and serves as the inspiration of XmasNet. Lastly, ResNet, can be considered the deepest network among the architectures investigated in this study, with 50 layers. Its complexity is enabled by the use of residual connections between layers, which learn a reference residual mapping, making the training of arbitrarily deep CNNs theoretically possible.

2.2. Conditional Random Fields as Recurrent Neural Networks

CRFs achieved state-of-the-art results in the image segmentation tasks, both in the traditional benchmark [47], as well as in application to medical image analysis, such as in PCa segmentation [48], weakly supervised segmentation of PCa [49], GS grading [50] or PCa detection [51].

CRF functioning is based on the notion of energy $\mathcal{E}(\cdot)$, i.e., the cost of assigning a label to a given pixel. A CRF is composed of two types of energy—namely, unary and pairwise—which must agree and can be described as:

$$\mathcal{E}(x) = \sum_i \Psi_u(x_i) + \sum_p \Psi_p(x_i, x_j). \tag{1}$$

The unary energy $\Psi_u(x_i)$ is the probability of a pixel i belonging to a given label x_i in this case extracted by a CNN. Conversely, $\Psi_p(x_i, x_j)$ corresponds to the pairwise energy. It measures the cost of assigning the labels x_i and x_j to pixels i and j simultaneously. It ensures image smoothness and consistency; pixels with similar properties should have similar labels. Thus, CRFs promote regions of homogeneous predictions. The pairwise energy is defined according to [52]:

$$\Psi_p(x_i, x_j) = \mu(x_i, x_j)\mathcal{K}(\mathbf{f}_i, \mathbf{f}_j), \tag{2}$$

where $\mu(x_i, x_j)$ is a label compatibility function and the Gaussian kernel $\mathcal{K}(\cdot, \cdot)$ applied on the feature vectors $\mathbf{f}_i$ and $\mathbf{f}_j$, with $w^{(m)}$ being linear combination weights:

$$\mathcal{K}(\mathbf{f}_i, \mathbf{f}_j) = \sum_{m=1}^{k} w^{(m)}\mathcal{K}^{(m)}(\mathbf{f}_i, \mathbf{f}_j). \tag{3}$$

In our case, the features consider the positions p_i and p_j as well as the intensity values I_i and I_j of the pixels in the image:

$$k(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right)}_{\text{appearance kernel}} + w^{(2)} \underbrace{\exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right)}_{\text{smoothness kernel}}, \tag{4}$$

with θ_α, θ_β and θ_γ being hyper-parameters controlling the importance of the hyper-voxel distance in the feature space (i.e., the appearance kernel promotes pixels with similar intensity values to be in the same class, while the smoothness kernel removes small isolated regions [52])—a stronger penalty is given if nearby pixels have different labels. In this version, the Potts model is used, i.e., $\mu(x_i, x_j) = [x_i \neq x_j]$ [31,52].

The training time for a CRF grows exponentially with respect to the number of input pixels N, even when considering approximate training methods, like Markov Chain Monte Carlo (MCMC), pseudo-likelihood or junction trees [31,53]. To solve this shortcoming, the Mean Field approximation (MFa) can be used [52]. MFa consists in approximating the distribution $P(\mathbf{X})$,—where $\mathbf{X}$ is the vector of the random variables $X_1, X_2, \ldots, X_N$ denoting the N pixels composing the image—by a simpler distribution $Q(\mathbf{X})$, which can be written as the product of independent marginal distributions: $Q(\mathbf{X}) = \prod_i Q_i(\mathbf{X}_i)$, subject to: $\sum_{x_i} Q_i(\mathbf{X}_i) = 1$.

The function $Q_i(x_i)$ can be defined such that is updated iteratively [52]. Bearing this in mind, an MFa-trained CRF still cannot be trained by means of the backpropagation, making its end-to-end integration with CNN infeasible, as the CNN and CRF need to be trained separately. Further refining MFa, the authors of [31] formalized the CRFs as Recurrent Neural Networks (CRF-RNN). This work redefines the MFa as a series of convolutional and recurrent layers. The convolutional layers perform the Gaussian operations on the features via learnable filters, while the recurrent layers behave as several iterations of the MFa method. This joint approach, by unrolling the CRF MFa inference steps, builds an end-to-end trainable feed-forward network composed of an initial CNN component performing feature extraction and a CRF-based probabilistic graphical model component for structured prediction, without the need for two separate training procedures. Since the two components can learn in the same environment, they can cooperate to achieve the best performance [31]. Thereby, with this implementation, a CRF-RNN network is limited to a batch size of 1 due to GPU memory constraints [31].

FC CRFs achieved outstanding performance in semantic segmentation [52,54] and remote sensing [55] when used as a postprocessing strategy. Interestingly, the results in [31] showed an evident competitive advantage of the joint end-to-end framework with respect to the offline application of CRFs as postprocessing method (disconnected from the CNN training). This might be attributed to the fact that during the backpropagation-enabled joint training, the CNN and CRF components cooperate to yield an optimized output by iteratively incorporating the CRF contribution. Relying on these experimental findings, we devised a hybrid end-to-end trainable model by introducing the CRF-RNN into XmasNet, along with three skip connections to merge the information from multiple layers. It is worth noting that we chose XmasNet as baseline since it was specifically designed for the PROSTATEx17 Challenge and is not a particularly deep architecture; therefore, it can serve as a suitable case study for evaluating the effective benefits achieved by the integration of the CRF-RNN module into CNNs.

3. Materials and Methods

This section presents the analyzed prostate MRI datasets, as well as the proposed end-to-end deep learning framework combining CRF-RNN and XmasNet.

3.1. Experimental Dataset: The PROSTATEx17 Dataset

This work considers the MRI dataset provided by the PROSTATEx Challenge 2017 [43] as part of the 2017 SPIE Medical Imaging Symposium [56], organized by the Society of Photographic Instrumentation Engineers (SPIE) and supported by the American Association of Physicists in Medicine (AAPM) and the National Cancer Institute (NCI). The aim of the PROSTATEx Challenge 2017 is to develop a quantitative diagnostic classification method of prostate lesions. This dataset was previously collected and curated by the Radboud University Medical Centre (Nijmegen, the Netherlands) in the Prostate MR Reference Center under the supervision of Prof. Jelle Barentsz [18].

The dataset contains mpMRI studies of 344 subjects. All studies include T2w, PDw, DCE, and DWI MRI sequences, acquired using the MAGNETOM Trio and Skyra 3T MRI scanner models from Siemens (Siemens Healthineers, Erlangen, Germany), without employing an endorectal coil. For more details on image acquisition, please refer to [43]. Although DCE imaging conveys relevant functional information, it carries the drawback of needing an external agent, typically via the injection of a Gadolinium-based

contrast [32,57]. The contrast may cause discomfort to the patient, with an increase in the risk of residual deposition in the human body [58] and without proof of an improvement in cancer detection quality [32]. In this study, we used non-contrast-enhanced MRI sequences only, since DWI—and especially ADC maps [59]—showed promising applications in the clinic. In a very recent study [60], similar cancer detection rates from biparametric MRI (bpMRI)—focusing on T2w and DWI—and contrast-enhanced mpMRI, particularly for Clinically Significant (CS) cases of PCa, were achieved. Examples of input T2w, PDw and ADC MR images are shown in Figures 1a, 1b and 1c, respectively.

Each MRI study was evaluated under the supervision of an expert radiologist that identified areas of suspicion. If an area was marked as likely for cancer, a biopsy was performed and then graded by a pathologist. If the biopsy results had a GS higher than 7, it was considered CS [18]. The ultimate goal of the PROSTATEx Challenge 2017 is to predict the clinical significance of a patient's lesion based on his MRI studies.

The whole cohort was divided into two sub-sets; each lesion's CS information was available in the training set (204 subjects) but not in the test set (140 subjects). Therefore, only the training set was considered in this study.

3.2. The Proposed End-to-End Solution Integrating CRF-RNN with CNNs for PCa Detection

The processing phases of the proposed end-to-end method are described in this section.

3.2.1. Data Preprocessing

Considering that the images were collected in different conditions (e.g., different scanners and acquisition configurations), an intermediate data preprocessing step was deemed necessary to ensure reliable data properties. The available images were characterized by having a different resolution in the 3D space (i.e., anisotropic voxels). Therefore, isotropic cubic interpolation was used on every image to achieve a resolution of 1.0 mm^3.

An image registration procedure was also necessary because several factors can contribute to making the individual studies not comparable (i.e, patient movement, different machine configurations) [61]. In image registration, a set of transformations τ is applied on one image (i.e., the moving image) so that its landmarks/features or pixel intensities overlap onto another reference (i.e., the fixed image), maximizing the matching of the pair. Only affine transformations (i.e., rotation, translation, and shearing) were considered. Mutual Information Criteria [62] were used to measure the quality of the registration. T2w was set as the fixed image, while PDw and ADC were used as moving images. The image interpolation and registration were performed by using the Diffusion Imaging Python (DIPy) package [63].

To extract the lesion information, a 64×64-pixel Region of Interest (RoI) was centered on the PCa coordinates. Z-score normalization was then employed to transform all the pixel values of each MRI slice to a common scale with zero mean and unitary standard deviation. The normalized RoI patches extracted from the T2w, ADC, and PDw MRI sequences were concatenated as a $64 \times 64 \times 3$ image to serve as the model inputs, as illustrated in Figure 1.

At the end of this procedure, considering the 204 subjects with CS annotations, 335 lesions were collected, corresponding to 20.5% of the lesions analyzed.

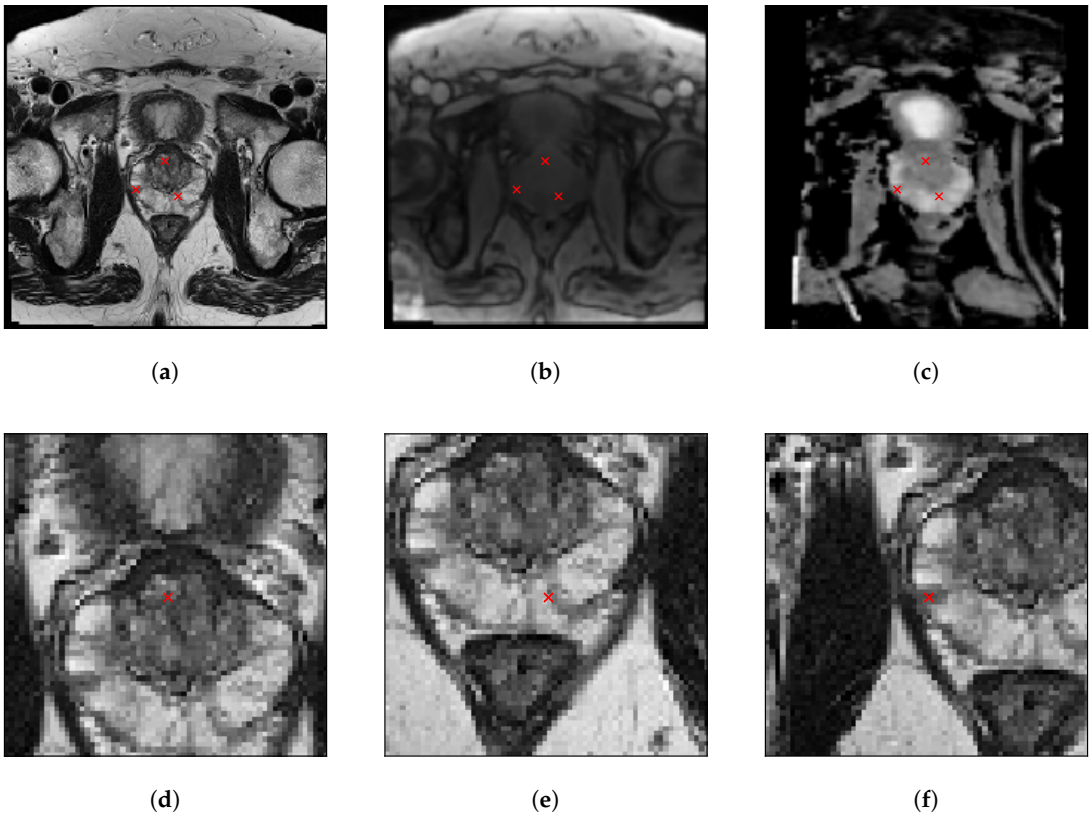

Figure 1. Example MR images from Patient #005: (**a**) T2w, (**b**) PDw, and (**c**) ADC sequences. The PCa lesions are highlighted (with red crosses) in the coordinates $(95, 92)$, $(103, 113)$, and $(78, 109)$, with slice index $z = 30$, after the registration against the T2w MRI sequence. The three lesions are separately displayed on the T2w slice in (**d**–**f**). In particular, these RoIs are centered and cropped on each lesion.

3.2.2. The CRF-XmasNet Architecture

We started from the XmasNet [33] architecture as baseline in our case study on the PROSTATEx17 dataset [43], since it achieved the second highest AUC in the PROSTATEx17 challenge despite its simplicity. Indeed, XmasNet can be seen as a parameter-efficient version of the VGG net [44] (i.e., XmasNet is less deep than VGG16) tailored to PCa detection in order to show the potential of deep learning in oncological imaging [33]. It is worth noting that the XmasNet model in the original paper exploited the ensemble of 20 individual XmasNet instances (which maximize the validation AUC on different combinations of the input MRI sequences) by using the weighted average of the predictions via a greedy bagging algorithm [33]. In our work, to keep full control of the CRF-RNN [31] introduction and to avoid the stochastic effect due to the network ensemble, we used only one XmasNet model to obtain the lesion's classification. In this manner, we can effectively assess the benefits provided by the end-to-end training of the proposed hybrid CRF-CNN approach.

Figure 2 schematizes the proposed CRF-XmasNet architecture. The network can be divided into three main components: (*i*) downsampling, based on convolution, Batch Normalization (BN), ReLU and Max Pooling operators; (*ii*) upsampling, using one-dimensional convolution and deconvolution operators; (*iii*) classification involving flatten and FC layers, along with ReLU and sigmoid activation factions. The details about these three components are provided in Tables 1–3, respectively. Importantly, a CRF-RNN [31] layer was integrated into the XmasNet architecture, utilizing and merging the features extracted by the convolutional portions as inputs (see Figure 2). The merging operation is performed via skip connections from multiple layers and summing-up the feature maps.

Additional modifications were proposed, aimed at improving the network effectiveness:

- three skip connections and two convolutional layers were added, as shown in Figure 2;
- dropout [64] was introduced between the FC layers and the sigmoid (with a dropout rate of 0.5);
- the number of parameters in the first and second FC layers was changed, from 1024 to 128 and 1024 to 256, respectively, because of performance constraints of the available computing power.

Skip connections were added as they allow for: (*i*) a simple method of merging information coming from several layers into the CRF; (*ii*) processing the input of higher-level features not present in the CRF layer output into the classifier component of the network; (*iii*) improving the training process [45], particularly during the early epochs, since in the backpropagation procedure some errors might be directly presented to the downsampling component.

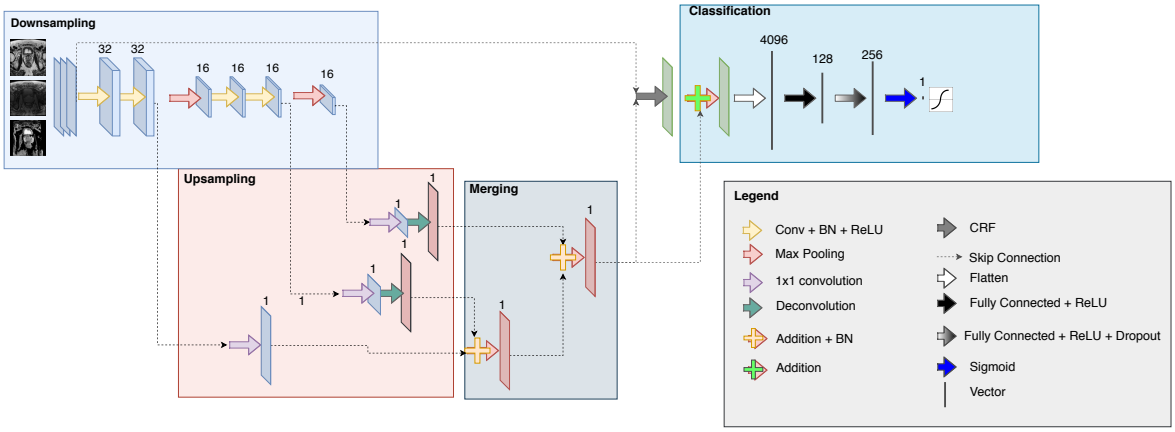

Figure 2. The proposed CRF-XmasNet architecture integrating CRFs [31] into the baseline XmasNet [33] as an end-to-end approach. This hybrid network, allowing for joint training via backpropagation, analyzes three non-contrast-enhanced mpMRI sequences (namely, T2w, T1w and ADC) and yields a prediction probability for each CS PCa case.More specifically, the whole architecture can be divided into three components: (*i*) downsampling, (*ii*) upsampling, and (*iii*) classification. In order to effectively merge the information from multiple layers into the CRF, three skip connections were added. The legend box shows the symbol notation and color semantics. The digits above the layers outputs represent the depth.

Table 1. Downsampling component network parameters.

Layer	Conv1	Conv2	MaxPool1	Conv3	Conv4	MaxPool2
Patch size/stride	$3 \times 3/1$	$3 \times 3/1$	$2 \times 2/2$	$3 \times 3/1$	$3 \times 3/1$	$2 \times 2/2$
Output size	$64 \times 64 \times 32$	$64 \times 64 \times 32$	$32 \times 32 \times 32$	$32 \times 32 \times 32$	$32 \times 32 \times 32$	$16 \times 16 \times 32$

Table 2. Upsampling component network parameters.

Layer	1d_Conv1	1d_Conv2	1d_Conv3	Deconv1	Deconv2
Input Layer	Conv2	Conv4	MaxPool2	Out_1d_Conv2	Out_1d_Conv3
Patch size/stride	$1 \times 1/1$	$1 \times 1/1$	$1 \times 1/1$	$2 \times 2 \times 2$	$2 \times 2/4$
Output size	$64 \times 64 \times 64$	$32 \times 32 \times 1$	$16 \times 16 \times 1$	$64 \times 64 \times 1$	$64 \times 64 \times 1$

Table 3. Classification component network parameters.

Layer	FC1	FC2	FC3
Output size	128×1	256×1	1×1

To evaluate the performance of the proposed CRF-XmasNet architecture, we also compared its performance against the XmasNet architecture in which a CRF is used as a postprocessing phase.

More specifically, the traditional XmasNet architecture is trained and, subsequently, CRFs are used to possibly improve the classification performance of the XmasNet architecture. While the work of Zheng et al. [31] showed that the use of CRFs as a postprocessing method (i.e., independent from the CNN training) results in poor performance when compared with the joint end-to-end framework, we believe that this analysis is important to strengthen the suitability of the proposed CRF-XmasNet architecture. In the remainder of the paper, we denote the XmasNet architecture, which uses CRFs as a postprocessing step, as XmasNet-CRF-postprocessing (XmasNet-CRFpp).

Along with the integration of CRFs into XmasNet, the proposed hybrid end-to-end approach was applied also to VGG16 and AlexNet with the aim of showing its suitability. In practice, CRF-RNNs can be integrated into any CNN: the downsampling component of the CRF-XmasNet, illustrated in Figure 2, can be substituted by the feature extraction sub-network of any architecture (i.e., the layers, before the FC layers that perform the classification, responsible for extracting features). In particular, the VGG16 and AlexNet architectures were transformed into their respective CRF-VGG16 and CRF-AlexNet versions by defining the downsampling component as the layers preceding the flattening operation of their original architectures. Indeed, VGG16 and AlexNet are overall more complex than the baseline XmasNet, which remains our benchmark since it was specifically designed for the PROSTATEx17 challenge [43].

3.2.3. Experimental Setup and Implementation Details

CNN performance might be particularly susceptible to two problems: (*i*) hyper-parameter settings and (*ii*) sensitivity to initialization values. Aiming at ensuring reliable and repeatable results, we tested several configurations and then trained multiple times.

First, three partitions were created: training, validation, and testing, with 60%, 20% and 20% of the original dataset, respectively. For each architecture, twenty hyper-parameter configurations were randomly created; every configuration was repeatedly trained 30 times for 350 epochs or until the loss score (Binary Cross Entropy, BCE) did not improve over 1×10^{-4} in the last 15 epochs. After the training, the model was evaluated with the test set. The configuration with the highest average AUROC value was considered the best of each given architecture. More specifically, the XmasNet-CRFpp architecture was trained through a two-step procedure: first, the XmasNet was trained; subsequently, using the weights of the best model, the mpMRI features were extracted to form an intermediary dataset. The CRF-RNN model was trained on this dataset, thus employing CRFs for postprocessing.

The training was performed with a batch size of 6 for the state-of-the-art CNN architectures and of 1 for the hybrid CRF-RNN/CNN approach. In more detail, for the CNN architectures that do not integrate CRFs, a batch size of 6 was used, as it is the batch size value that showed a suitable trade-off between training time and performance. On the other hand, the hybrid CRF-XmasNet and XmasNet-CRFpp architectures, where the CRF was integrated within the CNN (as an end-to-end and as a postprocessing, respectively), used a batch size of 1, which was selected to avoid reaching the memory limits of the GPU (as also suggested in [31]).

We employed a random sample algorithm, selecting one of several possible values for each hyper-parameter, as described in Table 4. Only high values for momentum m were considered in the grid search, based on good empirical results during prototyping when compared to lower values (e.g., $m < 0.9$), as well as with the support of the literature [65].

Table 4. List of the hyper-parameters considered in the grid search.

Parameter	Values
Optimizer	{SGD, Adam, RMSPROP}
Learning rate (l_r)	$\{1 \times 10^{-1}, 1 \times 10^{-2}, 1 \times 10^{-3}\ 1 \times 10^{-4}, 1 \times 10^{-5}, 1 \times 10^{-6}, 1 \times 10^{-7}\}$
Momentum (m)	{0.0, 0.9, 0.99, 0.999}
Decay (d)	{0.0, 0.1, 0.01, 0.001}
Nesterov (n)	{0, 1}
Amsgrad (a)	{0, 1}
CRF θ_α	{0.5, 1, 2, 3}
CRF θ_β	{0.5, 1, 2, 3}
CRF θ_γ	{0.5, 1, 2, 3}

The proposed framework was developed in Keras [66] with TensorFlow backend [67]. Two computational platforms were used: the prototyping was conducted with a laptop equipped with 8 GB of RAM, an Nvidia 840M GPU and an Intel i7-4510U 2.00 GHz CPU; the training was performed on a remote server with 8 GB of RAM, an Nvidia 1080ti GPU and an Intel i7-4790K 4.00 GHz CPU.

4. Results

This section presents and discusses the achieved experimental results. The best performing configuration of each architecture is shown in Table 5. These results suggest that the choice of the parameters is strictly related to the considered architecture, and it is something that must be taken into account before tackling the PCa classification problem. Table 6 reports the average values of loss and AUROC on the training and test sets, allowing for a comparative analysis of the different architectures considered in this study.

Table 5. Best performing configuration of each architecture. NA denotes Not Applicable.

Architecture	Optimizer	l_r	m	d	n	a	θ_α	θ_β	θ_γ
XmasNet	RMSPROP	1×10^{-5}	NA	0	NA	NA	NA	NA	NA
AlexNet	RMSPROP	1×10^{-5}	NA	0	NA	NA	NA	NA	NA
VGG16	Adam	1×10^{-4}	0.999	0	0	0	NA	NA	NA
ResNet	Adam	1×10^{-5}	0.99	0	0	0	NA	NA	NA
XmasNet-CRFpp	RMSPROP	1×10^{-5}	NA	0	NA	NA	1	1	0.5
CRF-XmasNet	RMSPROP	1×10^{-4}	0.99	0.1	NA	NA	1	1	0.5
CRF-AlexNet	RMSPROP	1×10^{-3}	0.99	0.01	NA	NA	2	3	2
CRF-VGG16	RMSPROP	1×10^{-4}	0	0	NA	NA	2	2	3

Table 6. Loss and AUROC values for each architecture. The average value (with standard deviation in parenthesis), obtained over 30 independent runs is shown.

Architecture	Loss Train	Loss Test	AUROC Train	AUROC Test
XmasNet	0.500 ± (0.023)	0.540 ± (0.043)	0.622 ± (0.054)	0.517 ± (0.101)
AlexNet	0.096 ± (0.032)	0.537 ± (0.023)	1.000 ± (0.000)	0.588 ± (0.051)
VGG16	0.474 ± (0.021)	0.481 ± (0.018)	0.729 ± (0.047)	0.707 ± (0.050)
ResNet	0.496 ± (0.020)	0.528 ± (0.023)	0.658 ± (0.065)	0.520 ± (0.100)
XmasNet-CRFpp	0.433 ± (0.082)	0.730 ± (0.242)	0.796 ± (0.069)	0.388 ± (0.303)
CRF-XmasNet	0.507 ± (0.077)	0.566 ± (0.097)	0.695 ± (0.126)	0.573 ± (0.191)
CRF-AlexNet	1.150 ± (0.366)	1.260 ± (0.392)	0.536 ± (0.191)	0.598 ± (0.169)
CRF-VGG16	0.644 ± (0.267)	0.732 ± (0.206)	0.796 ± (0.209)	0.615 ± (0.147)

According to these results, AlexNet is the best architecture when considering the values of the loss function on the training instances, but it severely overfits the training data, as shown when considering the test set. On the training set, the second-best performer is XmasNet-CRFpp, but also for this architecture a severe amount of overfitting can be noticed when one focuses on the loss function on

the test set. Still considering the training set, the third-best performer is VGG16, which also produces the lowest (i.e., the best) average loss function value on the test set. When comparing XmasNet and CRF-Xmasnet, it is worth noting that the two architectures produce similar values for the loss function on the training set, and they are the worst performers when compared against the other investigated architectures. When considering the values of the loss function produced by the networks that integrate CRFs into their basic architecture, one can notice that this combination achieves the worst (i.e., highest) values on the training set, but it allows us to reduce the amount of overfitting produced by some of the basic architectures. This is particularly evident when AlexNet and CRF-AlexNet are compared. In this case, CRF is actually working as a global regularizer.

More interesting experimental findings can be extracted when focusing on the AUROC values. In particular, AlexNet outperforms the remaining architectures on the training set, but due to the overfitting that affects the resulting model, its performance on the test set are significantly lower. More in detail, when considering the test set, VGG16 outperforms the other networks in terms of AUROC. When comparing the AUROC values obtained by XmasNet and CRF-XmasNet, the architecture with CRFs outperforms XmasNet on both the training and test instances. Additionally, the XmasNet-CRFpp architecture outperforms CRF-XmasNet and XmasNet on the training set, but its performance are the poorest on the test set. This behavior is aligned with the analysis performed by Zheng et al. [31], in which the use of CRFs as a postprocessing step was compared against an end-to-end approach. With regard to CRF-AlexNet and CRF-VGG16, different behaviors can be noticed. While the use of CRFs with AlexNet is beneficial for reducing overfitting, as well as for obtaining a model with an AUROC on the test set that is higher compared to the baseline architecture (i.e., AlexNet), the effect on VGG16 is different. In particular, the use of CRFs within VGG16 causes moderate overfitting, thus producing a final model that is not able to outperform VGG16 on the test set. All in all, considering the architectures that use CRFs, it is possible to observe that they, on average, perform comparably on the test set. This could be explained with the regularization effect obtained by using CRFs, and it is an aspect that deserves future investigation to better understand the outstanding peak performance, along with the high-variability of the results, obtained via CRFs. While the analysis of the average values does not show a particular advantage in considering CRFs for solving the task at hand, a more in-depth analysis shows an interesting phenomenon. In particular, Figure 3 displays the boxplots of the AUROC on the test set for the proposed CRF-XmasNet, and the other considered architectures. According to these figures, it is interesting to note that the best performing model was obtained by using CRFs in the end-to-end approach. Additionally, all the architectures that integrate CRFs yielded the best peak performance (i.e., the highest AUROC values). Nonetheless, CRF-RNN presents a considerably higher variability in terms of both AUROC and BCE when compared against the other architectures. Thus, an additional systematic investigation must be dedicated to a better understanding of this experimental evidence. Being able to identify the reasons that lead to this high-variability may allow us to modify the training process to prevent this behavior and guide the training process towards the identification of a model with performance that cannot be achieved by the existing network architectures. We hypothesize that the batch size of 1 might cause the high-variability of the CNNs' performance embedding CRF-RNNs. Unfortunately, due to the hardware limitations, we were not able to investigate the effect of this parameter on the performance of the network: the literature suggests that a small batch size value can result in a poor convergence of the model towards the global optima [68]. Focusing on the differences between XmasNet-CRFpp and CRF-RNN, one can notice that the former architecture presents an even greater variability and, moreover, it produces worse performance when compared with the proposed CRF-RNN architecture. This fact highlights the importance of using CRFs in the end-to-end model, thus exploiting CRFs' capabilities in leveraging feature relationships during the training process of the RNN.

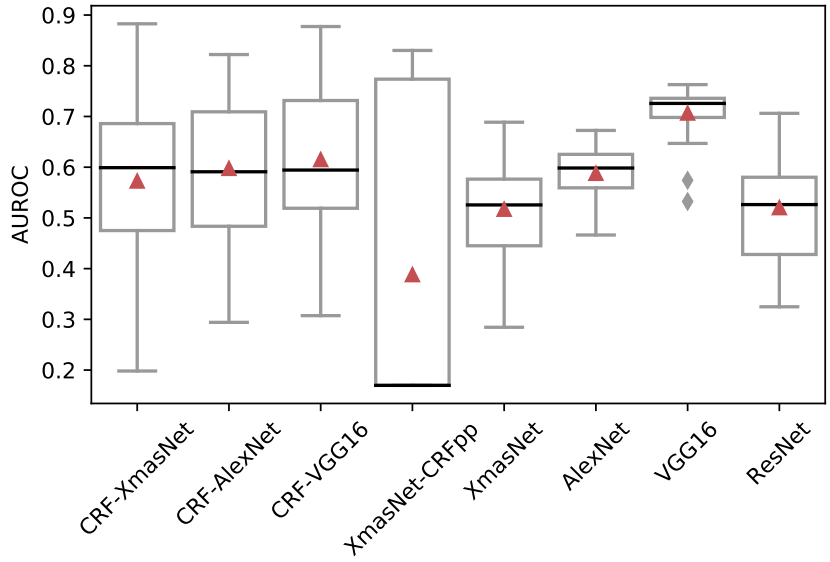

Figure 3. Boxplots of the AUROC obtained on the test set by the different architectures over 30 independent runs. Each boxplot shows a solid black line and a red triangle marker that denote the median and mean values, respectively.

From the analysis of the results, it is possible to draw some conclusions: focusing on the base architectures (i.e., without embedding the CRF-RNN), the very deep networks (i.e., VGG16 and AlexNet) are the best performers (both in terms of BCE and AUROC) when compared against the other state-of-the-art architectures (i.e., ResNet and XmasNet). The poor performance of ResNet may be justified by the reduced number of training images available for such a complex network. Lastly, there is an improvement between the CRF-XmasNet and XmasNet, as shown in the test sets (AUROC = 0.572 vs. 0.517, respectively). Considering the CRF-RNN component, CRF-XmasNet is characterized by a high-variability of AUROC and BCE, but it showed superior performance compared to XmasNet and XmasNet-CRFpp. CRF-AlexNet provides comparable performance with respect to the baseline AlexNet architecture, while embedding the CRF-RNN component into VGG16 results in poorer performance (in terms of test AUROC) than the baseline VGG16 architecture. The high-variability of CRF-XmasNet (approximately 1.89× and 3.82× higher than those of XmasNet and VGG16, respectively) deserves an in-depth investigation. We might argue that this variability causes the performance differences to not be statistically significant, as shown in Table 7, where the p-values of the Wilcoxon test on paired data (with a significance level $\alpha = 0.05$) from the test AUROC metrics are provided. On the other hand, being able to understand the source of this variability could help in guiding the search process towards highly-performing models that are characterized by AUROC values that cannot be achieved by relying on the other state-of-the-art architectures. The same investigation regards CRF-AlexNet and CRF-VGG16 that also present high-variability in terms of performance. While several alternatives were tested to reduce this high-variability (i.e., batch normalization, different weight initialization or hyperparameters), no significant improvement was achieved without affecting the performance of the resulting network. However, despite this drawback, we believe that the use of CRF has potential. In particular, focusing on CRF-XmasNet, even with the introduction of an early stopping criterion that limits the training epochs to 50, the performance improved while the training time was naturally reduced. More specifically, CRF-XmasNet required (on average) a training time of 249.8 s (with a standard deviation of 0.3 s), while XmasNet required 417.9 s (with a standard deviation of 32.7 s). A Wilcoxon test (with a significance level $\alpha = 0.05$) was executed to statistically assess these results. A p-value of 2.1×10^{-9} suggests that the difference (in terms of running time) between the two architectures is statistically significant.

Table 7. *P*-values obtained from the statistical validation procedure. The Wilcoxon rank-sum test for pairwise data comparison was used with the alternative hypothesis that the samples do not have equal medians of AUROC (test set values). A significance level of $\alpha = 0.05$ with a correction for multiple comparisons was used. **Boldface** indicates that the null hypothesis can be rejected.

	ResNet	CRF-XmasNet	CRF-AlexNet	CRF-VGG16	XmasNet-CRFpp	XmasNet	AlexNet
CRF-XmasNet	5.65×10^{-1}						
CRF-AlexNet	$\mathbf{1.24 \times 10^{-2}}$	9.12×10^{-1}					
CRF-VGG16	$\mathbf{9.88 \times 10^{-3}}$	4.35×10^{-1}	6.03×10^{-1}				
XmasNet-CRFpp	3.85×10^{-2}	$\mathbf{1.61 \times 10^{-2}}$	$\mathbf{7.78 \times 10^{-3}}$	$\mathbf{1.68 \times 10^{-3}}$			
XmasNet	9.05×10^{-1}	5.65×10^{-1}	9.12×10^{-1}	4.35×10^{-1}	3.32×10^{-2}		
AlexNet	$\mathbf{1.66 \times 10^{-2}}$	6.89×10^{-1}	6.82×10^{-1}	2.46×10^{-1}	$\mathbf{3.0 \times 10^{-4}}$	$\mathbf{1.66 \times 10^{-2}}$	
VGG16	$\mathbf{1.45 \times 10^{-8}}$	$\mathbf{2.56 \times 10^{-3}}$	3.26×10^{-2}	$\mathbf{1.10 \times 10^{-2}}$	$\mathbf{2.6 \times 10^{-4}}$	$\mathbf{5.55 \times 10^{-9}}$	$\mathbf{2.04 \times 10^{-8}}$

Lastly, Figure 4 shows that, although CRF-XmasNet reveals unstable performance, it also achieved top-of-the-class performance when compared against the best competitor, VGG16. In particular:

- in 8 out of 30 runs, the test AUROC of the CRF-XmasNet was higher than the best obtained with XmasNet;
- the top 6 best performing runs, considering CRF-XmasNet and XmasNet, were achieved by CRF- XmasNet;
- 19 of the 30 CRF-XmasNet runs obtained a performance higher than their average value;
- 22 of the CRF-XmasNet runs outperformed the XmasNet average value.

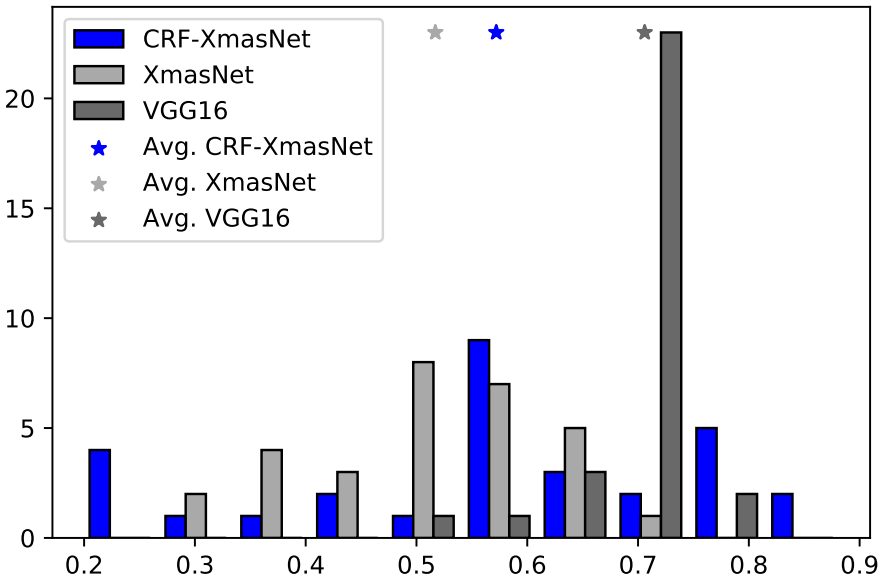

Figure 4. Bar graph with 10 groups that represent the occurrences of the test AUROC regarding XmasNet, CRF-XmasNet, and VGG16 architectures observed over 30 runs. The blue, light gray, and dark gray bars refer to the CRF-XmasNet, XmasNet and VGG16 architectures, respectively. The star markers of the same color lie on the average ROC values for the three architectures.

Overall, the proposed approach provides satisfactory performance when compared against XmasNet, a state-of-the-art CNN architecture specifically designed for dealing with PCa mpMRI data. These results suggest the suitability of integrating CRF-RNN within XmasNet. While experimental results showed very good performance achieved by VGG16 and AlexNet, these results could be related to the reduced size of the PROSTATEx17 dataset [69]. That is, considering a dataset with thousands of MRI studies, XmasNet might outperform the other competitors because it is specifically designed for extracting salient features from MRI data. Therefore, we believe that CRF-XmasNet provides an important contribution for practitioners in the medical imaging field, as it shows how the integration of CRFs into XmasNet outperforms the baseline XmasNet architecture.

5. Discussion and Conclusions

In this work, the potential of integrating the CRF-RNN mechanism into a CNN architecture is presented, not only for MIA applications but also for the Image Classification field in general-purpose Computer Vision. This joint approach is built upon previous work that has shown that the combination of CRFs and CNNs in a hybrid end-to-end model can achieve promising performance across several benchmark datasets in image segmentation tasks [31,47,49,51,52,70]. The proposed CRF-XmasNet architecture leads to an interesting improvement over its baseline architecture (XmasNet [33]), and its best performance is comparable with the one obtained with deeper neural architectures, namely: AlexNet [34], VGG16 [44], and ResNet [45]. Additionally, our work showed that the use of CRFs as a postprocessing method is not suitable for the classification problem taken into account. This result corroborates the analysis reported in [31]. CRF-RNN can achieve competitive performance by also reducing the training time when compared against the baseline architecture. Despite these advantages, the integration of CRFs produces results characterized by a higher variability when compared against the other considered architectures. This phenomenon was observed also when the CRF-RNN component was integrated into AlexNet and VGG16. In this case, the two resulting architectures (i.e., CRF-AlexNet and CRF-VGG16) were characterized by high-variability of performance on both training and test sets.

The amount of homogeneous and well-prepared datasets represents an important challenge in biomedical imaging [24]. As a matter of fact, Deep Learning research has been recently focusing on issues related to medical imaging datasets with limited sample size, achieving promising performance by means of weakly-/semi-supervised schemes [49,71] as well as Generative Adversarial Network (GAN)-based data augmentation [72,73]. Moreover, methods tailored to each particular clinical application should be devised, such as for improving the model generalization abilities even in the case of small datasets collected from multiple institutions [15].

Given the common ground that the Machine Learning and Image Classification fields share, more promising and robust performance may be achieved by the further integration of CRFs into CNNs. For instance, CNN architecture tuning [74] might reduce the variability encountered in the experiments involving CRF-XmasNet and the other CNNs embedding CRFs. This contribution can open additional research directions aimed at investigating the performance variability of the CRF-RNNs when integrated into CNNs as an end-to-end approach, thus allowing for their use in a clinical environment.

In conclusion, the proposed end-to-end PCa detection approach might be used as a CAD, before the conventional mpMRI interpretation by experienced radiologists, aiming at increasing sensitivity and reducing operator dependence upon multiple readers [6]. To improve CS PCa classification performance, the combination with metabolic imaging might provide complementary clinical insights into tumor responses to oncological therapies [75]. Novel nuclear medicine tracers for Positron Emission Tomography (PET) [76] and hyperpolarized carbon-13 (^{13}C) and sodium (^{23}Na) MRI [77] can considerably improve the specificity for evaluating PCa with respect to conventional imaging, by understanding the pyruvate conversion to lactate for estimating the cancer grade [78]. From a computational perspective, novel solutions must be devised to combine multi-modal imaging data [79]. In the case of DNNs, a topology explicitly designed for information exchange—between sub-networks processing the data from a single modality—through cross-connections, such as in the case of cross-modal CNNs (X-CNNs) [80], might be suitable for combining multi-modal imaging data.

Author Contributions: Conceptualization, M.C. and L.R.; Investigation, P.L., M.C. and L.R.; Methodology, P.L., M.C. and L.R.; Software, P.L.; Supervision, M.C. and L.R.; Validation, I.G., E.S. and L.R.; Writing—original draft, P.L., M.C. and L.R.; Writing—review & editing, I.G. and E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by projects UID/MULTI/00308/2019 and by the European Regional Development Fund through the COMPETE 2020 Programme, FCT—Portuguese Foundation for Science and Technology and Regional Operational Program of the Center Region (CENTRO2020) within project MAnAGER

(POCI-01-0145-FEDER-028040). This work was also partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia) under project DSAIPA/DS/0022/2018 (GADgET) and by the financial support from the Slovenian Research Agency (research core funding No. P5-0410). This work was partially supported by The Mark Foundation for Cancer Research and Cancer Research UK Cambridge Centre [C9685/A25177]. Additional support has been provided by the National Institute of Health Research (NIHR) Cambridge Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA Cancer J. Clin.* **2019**, *69*, 7–34. [CrossRef] [PubMed]
2. 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.* **2018**, *68*, 394–424. [CrossRef] [PubMed]
3. Barlow, H.; Mao, S.; Khushi, M. Predicting high-risk prostate cancer using machine learning methods. *Data* **2019**, *4*, 129. [CrossRef]
4. Turkbey, B.; Brown, A.M.; Sankineni, S.; Wood, B.J.; Pinto, P.A.; Choyke, P.L. Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer. *CA Cancer J. Clin.* **2016**, *66*, 326–336. [CrossRef]
5. Yadav, S.S.; Stockert, J.A.; Hackert, V.; Yadav, K.K.; Tewari, A.K. Intratumor heterogeneity in prostate cancer. *Urol. Oncol.* **2018**, *36*, 349–360. [CrossRef]
6. Greer, M.D.; Lay, N.; Shih, J.H.; Barrett, T.; Bittencourt, L.K.; Borofsky, S.; Kabakus, I.; Law, Y.M.; Marko, J.; Shebel, H.; et al. Computer-aided diagnosis prior to conventional interpretation of prostate mpMRI: an international multi-reader study. *Eur. Radiol.* **2018**, *28*, 4407–4417. [CrossRef]
7. Stoyanova, R.; Takhar, M.; Tschudi, Y.; Ford, J.C.; Solórzano, G.; Erho, N.; Balagurunathan, Y.; Punnen, S.; Davicioni, E.; Gillies, R.J.; et al. Prostate Cancer Radiomics Promise Radiogenomics. *Transl. Cancer Res.* **2016**, *5*, 432. [CrossRef]
8. Choi, Y.J.; Kim, J.K.; Kim, N.; Kim, K.W.; Choi, E.K.; Cho, K.S. Functional MR imaging of prostate cancer. *Radiographics* **2007**, *27*, 63–75. [CrossRef]
9. Hegde, J.V.; Mulkern, R.V.; Panych, L.P.; Fennessy, F.M.; Fedorov, A.; Maier, S.E.; Tempany, C.M. Multiparametric MRI of prostate cancer: An update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *J. Magn. Reson. Imaging* **2013**, *37*, 1035–1054. [CrossRef]
10. Lemaître, G.; Martí, R.; Freixenet, J.; Vilanova, J.C.; Walker, P.M.; Meriaudeau, F. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. *Comput. Biol. Med.* **2015**, *60*, 8–31. [CrossRef]
11. Haider, M.A.; Van Der Kwast, T.H.; Tanguay, J.; Evans, A.J.; Hashmi, A.T.; Lockwood, G.; Trachtenberg, J. Combined T2-weighted and diffusion-weighted MRI for localization of prostate cancer. *Am. J. Roentgenol.* **2007**, *189*, 323–328. [CrossRef] [PubMed]
12. Fabijańska, A. A novel approach for quantification of time–intensity curves in a DCE-MRI image series with an application to prostate cancer. *Comput. Biol. Med.* **2016**, *73*, 119–130. [CrossRef] [PubMed]
13. Orczyk, C.; Villers, A.; Rusinek, H.; Lepennec, V.; Bazille, C.; Giganti, F.; Mikheev, A.; Bernaudin, M.; Emberton, M.; Fohlen, A.; et al. Prostate cancer heterogeneity: texture analysis score based on multiple magnetic resonance imaging sequences for detection, stratification and selection of lesions at time of biopsy. *BJU Int.* **2019**, *124*, 76–86. [CrossRef] [PubMed]
14. Rundo, L.; Militello, C.; Russo, G.; Garufi, A.; Vitabile, S.; Gilardi, M.C.; Mauri, G. Automated prostate gland segmentation based on an unsupervised fuzzy c-means clustering technique using multispectral T1w and T2w MR imaging. *Information* **2017**, *8*, 49. [CrossRef]
15. Rundo, L.; Han, C.; Nagano, Y.; Zhang, J.; Hataya, R.; Militello, C.; Tangherloni, A.; Nobile, M.S.; Ferretti, C.; Besozzi, D.; et al. USE-Net: Incorporating Squeeze-and-Excitation blocks into U-Net for prostate zonal segmentation of multi-institutional MRI datasets. *Neurocomputing* **2019**, *365*, 31–43. [CrossRef]

16. Wibmer, A.; Hricak, H.; Gondo, T.; Matsumoto, K.; Veeraraghavan, H.; Fehr, D.; Zheng, J.; Goldman, D.; Moskowitz, C.; Fine, S.W.; et al. Haralick Texture Analysis of prostate MRI: Utility for differentiating non-cancerous prostate from prostate cancer and differentiating prostate cancers with different Gleason Scores. *Eur. Radiol.* **2015**, *25*, 2840–2850. [CrossRef]
17. Park, J.; Rho, M.J.; Park. Y.H.; Jung, C.K.; Chong, Y.; Kim, C.S.; Go, H.; Jeon, S.S.; Kang, M.; Lee, H.J.; et al. PROMISE CLIP Project: A Retrospective, Multicenter Study for Prostate Cancer that Integrates Clinical, Imaging and Pathology Data. *Appl. Sci.* **2019**, *9*, 2982. [CrossRef]
18. Litjens, G.; Debats, O.; Barentsz, J.; Karssemeijer, N.; Huisman, H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans. Med. Imaging* **2014**, *33*, 1083–1092. [CrossRef]
19. Villeirs, G.M.; De Meerleer, G.O. Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning. *Eur. J. Radiol.* **2007**, *63*, 361–368. [CrossRef]
20. Quon, J.S.; Moosavi, B.; Khanna, M.; Flood, T.A.; Lim, C.S.; Schieda, N. False positive and false negative diagnoses of prostate cancer at multi-parametric prostate MRI in active surveillance. *Insights Imaging* **2015**, *6*, 449–463. [CrossRef]
21. Mangrum, W.; Christianson, K.; Duncan, S.; Hoang, P.; Song, A.; Merkle, E. *Duke Review of MRI Principles*, 1st ed.; Case Review Series; Elsevier: Maryland Heights, MI, USA, 2012.
22. Tofts, P.S. T1-weighted DCE imaging concepts: Modelling, acquisition and analysis. *Signal* **2010**, *500*, 400.
23. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Engin.* **2017**, *19*, 221–248. [CrossRef] [PubMed]
24. Ravì, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.-Z. Deep Learn. Health Informatics. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 4–21. [CrossRef] [PubMed]
25. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
26. Gonçalves, I.; Silva, S.; Fonseca, C.M. Semantic Learning Machine: A Feedforward Neural Network Construction Algorithm Inspired by Geometric Semantic Genetic Programming. In *Progress in Artificial Intelligence*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9273, pp. 280–285. [CrossRef]
27. Gonçalves, I. An Exploration of Generalization and Overfitting in Genetic Programming: Standard and Geometric Semantic Approaches. Ph.D. Thesis, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal, 2017.
28. Jagusch, J.B.; Gonçalves, I.; Castelli, M. Neuroevolution under Unimodal Error Landscapes: An Exploration of the Semantic Learning Machine Algorithm. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '18), Kyoto, Japan, 15–19 July 2018; ACM: New York, NY, USA, 2018. [CrossRef]
29. Lapa, P.; Gonçalves, I.; Rundo, L.; Castelli, M. Semantic learning machine improves the CNN-based detection of prostate cancer in non-contrast-enhanced MRI. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO) Companion, Prague, Czech Republic, 13–17 July 2019; ACM: Hoboken, NJ, USA, 2019; pp. 1837–1845. [CrossRef]
30. Lapa, P.; Gonçalves, I.; Rundo, L.; Castelli, M. Enhancing classification performance of convolutional neural networks for prostate cancer detection on magnetic resonance images: A study with the semantic learning machine. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO '19), Prague, Czech Republic, 13–17 July 2019; ACM: New York, NY, USA, 2019; pp. 381–382. [CrossRef]
31. Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Vineet, V.; Su, Z.; Du, D.; Huang, C.; Torr, P.H.S. Conditional random fields as recurrent neural networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 1529–1537. [CrossRef]
32. Junker, D.; Steinkohl, F.; Fritz, V.; Bektic, J.; Tokas, T.; Aigner, F.; Herrmann, T.R.; Rieger, M.; Nagele, U. Comparison of multiparametric and biparametric MRI of the prostate: Are gadolinium-based contrast agents needed for routine examinations? *World J. Urol.* **2018**, 1–9. [CrossRef] [PubMed]
33. Liu, S.; Zheng, H.; Feng, Y.; Li, W. Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. *Proc. SPIE* **2017**, *10134*, 1013428. [CrossRef]
34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv. Neural Inf. Process. Syst.* **2012**. [CrossRef]

35. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: http://www.deeplearningbook.org (accessed on 16 December 2019).
36. Wang, X.; Yang, W.; Weinreb, J.; Han, J.; Li, Q.; Kong, X.; Yan, Y.; Ke, Z.; Luo, B.; Liu, T.; et al. Searching for prostate cancer by fully automated magnetic resonance imaging classification: Deep learning versus non-deep learning. *Sci. Rep.* **2017**, *7*, 15415. [CrossRef]
37. Ampeliotis, D.; Antonakoudi, A.; Berberidis, K.; Psarakis, E.; Kounoudes, A. A computer-aided system for the detection of prostate cancer based on magnetic resonance image analysis. In Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP), St Julians, Malta, 12–14 March 2008; pp. 1372–1377. [CrossRef]
38. Guo, Y.; Gao, Y.; Shen, D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. *IEEE Trans. Med. Imaging* **2016**, *35*, 1077–1089. [CrossRef]
39. Fehr, D.; Veeraraghavan, H.; Wibmer, A.; Gondo, T.; Matsumoto, K.; Vargas, H.A.; Sala, E.; Hricak, H.; Deasy, J.O. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E6265–E6273. [CrossRef]
40. Bhattacharjee, S.; Park, H.G.; Kim, C.H.; Prakash, D.; Madusanka, N.; So, J.H.; Cho, N.H.; Choi, H.K. Quantitative analysis of benign and malignant tumors in histopathology: Predicting prostate cancer grading using SVM. *Appl. Sci.* **2019**, *9*, 2969. [CrossRef]
41. Jung, W.; Park, S.; Jung, K.H.; Hwang, S. Prostate cancer segmentation using manifold mixup U-Net. In Proceedings of the Medical Imaging with Deep Learning (MIDL), London, UK, 8–10 July 2019.
42. Ing, N.; Ma, Z.; Li, J.; Salemi, H.; Arnold, C.; Knudsen, B.; Gertych, A. Semantic segmentation for prostate cancer grading by convolutional neural networks. Medical Imaging 2018: Digital Pathology. *Proc. SPIE* **2018**, *10581*, 105811B. [CrossRef]
43. Armato, S.G.; Huisman, H.; Druk, K.; Hadjiiski, L.; Kirby, J.S.; Petrick, N.; Redmond, G.; Giger, M.L.; Cha, K.; Mamonov, A.; et al. PROSTATEx Challenges for computerized classification of prostate lesions from multiparametric magnetic resonance images. *J. Med. Imaging* **2018**, *5*, 044501. [CrossRef] [PubMed]
44. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [CrossRef]
46. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
47. Arnab, A.; Zheng, S.; Jayasumana, S.; Romera-Paredes, B.; Larsson, M.; Kirillov, A.; Savchynskyy, B.; Rother, C.; Kahl, F.; Torr, P.H. Conditional random fields meet deep neural networks for semantic segmentation: Combining probabilistic graphical models with deep learning for structured prediction. *IEEE Signal Process. Mag.* **2018**, *35*, 37–52. [CrossRef]
48. Artan, Y.; Langer, D.L.; Haider, M.A.; Van der Kwast, T.H.; Evans, A.J.; Wernick, M.N.; Yetik, I.S. Prostate cancer segmentation with multispectral MRI using cost-sensitive conditional random fields. In Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009; pp. 278–281. [CrossRef]
49. Kervadec, H.; Dolz, J.; Tang, M.; Granger, E.; Boykov, Y.; Ayed, I.B. Constrained-CNN losses for weakly supervised segmentation. *Med. Image Anal.* **2019**, *54*, 88–99. [CrossRef]
50. Jacobs, J.G.; Panagiotaki, E.; Alexander, D.C. Gleason Grading of Prostate Tumours with Max-Margin Conditional Random Fields; In *Machine Learning in Medical Imaging (MLMI), Proceedings of the 5th International Workshop, MLMI 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, 14 September 2014*; Wu, G., Zhang, D., Zhou, L., Eds.; LNCS; Springer International Publishing: Cham, Switzerland, 2014; Volume 8679, pp. 85–92. [CrossRef]
51. Monaco, J.P.; Tomaszewski, J.E.; Feldman, M.D.; Hagemann, I.; Moradi, M.; Mousavi, P.; Boag, A.; Davidson, C.; Abolmaesumi, P.; Madabhushi, A. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. *Med. Image Anal.* **2010**, *14*, 617–629. [CrossRef]

52. Krähenbühl, P.; Koltun, V. Efficient inference in fully connected CRFs with Gaussian edge potentials. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Granada, Spain, 12–14 December 2011; pp. 109–117.
53. Sutton, C.; McCallum, A. An Introduction to Conditional Random Fields. *Found. Trends® Mach. Learn.* **2012**, *4*, 267–373. [CrossRef]
54. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
55. Shrestha, S.; Vanneschi, L. Improved fully convolutional network with conditional random fields for building extraction. *Remote Sens.* **2018**, *10*, 1135. [CrossRef]
56. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Arch. (TCIA): Maintaining and operating a public information repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef] [PubMed]
57. Rundo, L.; Militello, C.; Vitabile, S.; Casarino, C.; Russo, G.; Midiri, M.; Gilardi, M.C. Combining split-and-merge and multi-seed region growing algorithms for uterine fibroid segmentation in MRgFUS treatments. *Med. Biol. Eng. Comput.* **2016**, *54*, 1071–1084. [CrossRef] [PubMed]
58. Gulani, V.; Calamante, F.; Shellock, F.G.; Kanal, E.; Reeder, S.B.; International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: Summary of evidence and recommendations. *Lancet Neurol.* **2017**, *16*, 564–570. [CrossRef]
59. Barrett, T.; Lawrence, E.M.; Priest, A.N.; Warren, A.Y.; Gnanapragasam, V.J.; Gallagher, F.A.; Sala, E. Repeatability of diffusion-weighted MRI of the prostate using whole lesion ADC values, skew and histogram analysis. *Eur. J. Radiol.* **2019**, *110*, 22–29. [CrossRef]
60. Sherrer, R.L.; Glaser, Z.A.; Gordetsky, J.B.; Nix, J.W.; Porter, K.K.; Rais-Bahrami, S. Comparison of biparametric MRI to full multiparametric MRI for detection of clinically significant prostate cancer. *Prostate Cancer Prostatic Dis.* **2019**, *22*, 331. [CrossRef]
61. Rundo, L.; Tangherloni, A.; Militello, C.; Gilardi, M.C.; Mauri, G. Multimodal medical image registration using particle swarm optimization: A review. In Proceedings of the Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–8. [CrossRef]
62. Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; Suetens, P. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* **1997**, *16*, 187–198. [CrossRef]
63. Garyfallidis, E.; Brett, M.; Amirbekian, B.; Rokem, A.; Van Der Walt, S.; Descoteaux, M.; Nimmo-Smith, I. DIPy, a library for the analysis of diffusion MRI data. *Front. Neuroinform.* **2014**, *8*, 8. [CrossRef]
64. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
65. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
66. Chollet, F. Keras. Available online: https://keras.io (accessed on 16 December 2019).
67. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado G.S.; Davis A.; Dean J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 16 December 2019).
68. Smith, S.L.; Kindermans, P.J.; Ying, C.; Le, Q.V. Don't decay the learning rate, increase the batch size. *arXiv* **2018**, arXiv:1711.00489.
69. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. *arXiv* **2017**, arXiv:1611.03530.
70. Monteiro, M.; Figueiredo, M.A.T.; Oliveira, A.L. Conditional random fields as recurrent neural networks for 3D medical imaging segmentation. *arXiv* **2018**, arXiv:1807.07464.
71. Cheplygina, V.; de Bruijne, M.; Pluim, J.P. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* **2019**, *54*, 280–296. [CrossRef] [PubMed]
72. Han, C.; Rundo, L.; Araki, R.; Nagano, Y.; Furukawa, Y.; Mauri, G.; Nakayama, H.; Hayashi, H. Combining Noise-to-Image and Image-to-Image GANs: Brain MR Image Augmentation for Tumor Detection. *IEEE Access* **2019**, *7*, 156966–156977. [CrossRef]

73. Han, C.; Kitamura, Y.; Kudo, A.; Ichinose, A.; Rundo, L.; Furukawa, Y.; Umemoto, K.; Li, Y.; Nakayama, H. Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. In Proceedings of the International Conference on 3D Vision (3DV), Québec City, QC, Canada, 16–19 September 2019; pp. 729–737. [CrossRef]
74. Janke, J.; Castelli, M.; Popovič, A. Analysis of the proficiency of fully connected neural networks in the process of classifying digital images benchmark of different classification algorithms on high-level image features from convolutional layers. *Expert Syst. Appl.* **2019**. [CrossRef]
75. Brindle, K. New approaches for imaging tumour responses to treatment. *Nat. Rev. Cancer* **2008**, *8*, 94. [CrossRef]
76. Lindenberg, L.; Choyke, P.; Dahut, W. Prostate cancer imaging with novel PET tracers. *Curr. Urol. Rep.* **2016**, *17*, 18. [CrossRef]
77. Barrett, T.; Riemer, F.; McLean, M.A.; Kaggie, J.D.; Robb, F.; Warren, A.Y.; Graves, M.J.; Gallagher, F.A. Molecular imaging of the prostate: Comparing total sodium concentration quantification in prostate cancer and normal tissue using dedicated 13C and 23Na endorectal coils. *J. Magn. Reson. Imaging* **2019**. [CrossRef]
78. Granlund, K.L.; Tee, S.S.; Vargas, H.A.; Lyashchenko, S.K.; Reznik, E.; Fine, S.; Laudone, V.; Eastham, J.A.; Touijer, K.A.; Reuter, V.E.; et al. Hyperpolarized MRI of human prostate cancer reveals increased lactate with tumor grade driven by Monocarboxylate Transporter 1. *Cell Metab.* **2019**. [CrossRef]
79. Rundo, L.; Stefano, A.; Militello, C.; Russo, G.; Sabini, M.G.; D'Arrigo, C.; Marletta, F.; Ippolito, M.; Mauri, G.; Vitabile, S.; et al. A fully automatic approach for multimodal PET and MR image segmentation in Gamma Knife treatment planning. *Comput. Methods Programs Biomed.* **2017**, *144*, 77–96. [CrossRef] [PubMed]
80. Veličković, P.; Wang, D.; Lane, N.D.; Liò, P. X-CNN: Cross-modal convolutional neural networks for sparse datasets. In Proceedings of the Symposium Series on Computational Intelligence (SSCI), Athens, Greece, 6–9 December 2016; pp. 1–8. [CrossRef]