

Using CNN With Handcrafted Features for Prostate Cancer Classification

1st Yimo Liu
School of Software Engineering
Tongji University
Shanghai, China
liuyimocn@gmail.com

2nd Di Bu
School of Software Engineering
Tongji University
Shanghai, China
budi20000427@126.com

3rd Guokai Zhang
School of Software Engineering
Tongji University
Shanghai, China
zhang.guokai@tongji.edu.cn

4th Ye Luo
School of Software Engineering
Tongji University
Shanghai, China
yelu@tongji.edu.cn

5th Jianwei Lu
School of Software Engineering
Tongji University
Shanghai, China
jwlu33@tongji.edu.cn

6th Weigang Wang*
Department of Radiology
Shanghai Fire Corps Hospital
Shanghai, China
wangweigang211@163.com

7th Binghui Zhao*
Department of Radiology
Shanghai Tenth People's Hospital
Shanghai, China
binghui Zhao@163.com

Abstract—Prostate cancer has been a leading cause of death among males for a long time. Currently, with the help of computer-aided detection systems, prostate cancer can be detected in a relatively early stage, thus improving the patients' survival rate. In this paper, we propose a computer-aided system based on deep learning method to help classify prostate cancer. Our model combines both convolutional neural network (CNN) extracted features and handcrafted features. In our model, the input data is sent into two subnets. One is a modified ResNet with an improved spatial transformer (ST) for high dimension feature extraction. The other subnet extracts three handcrafted features and processes them with a simple CNN. After those two subnets, the output features of the two subnets are concatenated and then sent into the final classifier for prostate cancer classification. Experimental results show that our model achieves an accuracy of 0.947, which is better than other state-of-the-art methods.

Index Terms—Prostate Cancer, Magnetic Resonance Imaging, Convolutional Neural Network, Handcrafted Feature

I. INTRODUCTION

Prostate cancer has been around as a cause of death since ancient times [1]. It's the most prevalent cancer and the third leading cause of death among males in the western countries [2]. At the clinical level, Magnetic Resonance Imaging (MRI) is introduced to the area of prostate cancer detection. As are showed in Fig.1, the T2-weighted images (T2WI) in the first row are all non-cancer ones while those in the second and third rows are diagnosed as prostate cancer. Those images provide the ability to acquire both anatomical and functional information, thus leading to a higher accuracy of prostate cancer detection [3].

With the development of computer technology, computer-aided detection (CAD) systems for prostate detection can

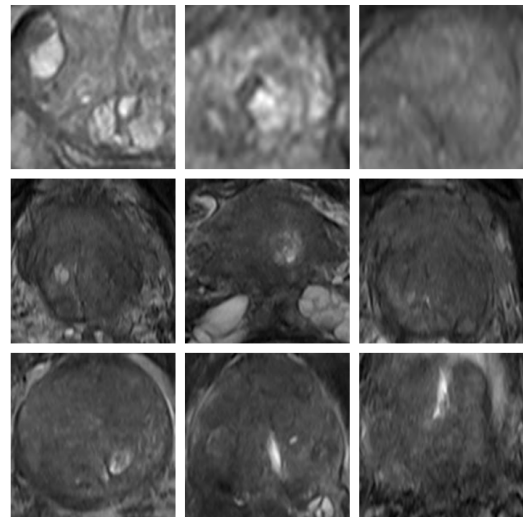


Fig. 1. Some Examples of T2WI, the images in the first row are all be non-cancer ones while those in the second and third rows are diagnosed as prostate cancer.

detect a suspicious lesion from processed magnetic resonance images with high efficiency. The early methods that use medical images to develop a computer-aided system were becoming increasingly popular. To form those systems, a crucial step is to extract various handcrafted features of images. However, due to the empirical definition of those handcraft features, such methods can't have a satisfactory performance when it comes to some more specific and complex issues like medical image

processing. Recently, with the great success of deep learning in the computer vision field, it has offered a potential and effective solution to solve that problem.

Compared to methods based on handcrafted feature extraction, CNN based methods can depict the features of images with more sensitivity and achieve a higher accuracy eventually. The performance of a deep learning method depends on the architecture of the CNN model. There came those general classification architectures like GoogLeNet [4] and ResNet [5]. However, those methods work perfectly on natural image classification but their performance on medical images is relatively less satisfying, showing that there exists great feasibility for us to improve them. To do this, other novel architectures [6] especially for medical image classification, arouse rapidly. Also, researchers tried to find other ways besides creating new architectures to enhance the performance, like modifying those classical models with some new units, or fine-tuning pre-trained models. Those approaches are commonly seen in the field of medical image classification including prostate cancer detection [7].

In our survey of the previous methods, we find that most of them have the main disadvantage of focusing only on one side of the issue. They either improve the existing algorithms for handcrafted features or try to construct new CNN architectures. But those two sides can likely both contribute to the improvement. Therefore, we propose a novel deep learning model for prostate cancer detection that consists of two subnets, a complex CNN for high dimension feature extraction and a simple CNN for handcrafted feature process. In this way, we manage to combine both CNN extracted features and handcrafted features. The two parts complement each other so well that we achieved a satisfying result. The main contributions of our work could be concluded as follows:

- (1) A model combining both handcrafted features and CNN extracted features for prostate cancer detection is proposed.
- (2) An improved ST module is utilized so that the model can learn the features with more transformation.
- (3) The experimental results validate that our method has better performance than other state-of-the-art methods.

II. RELATED WORK

A. Handcrafted Feature based Methods

In the past decades, designing efficient and robust image features was the primary goal of image classification tasks. These hand-crafted features describe the image from different perspectives. For example, Nketiah *et al.* [8] proposed a model by extracting the T2WI textural features for the prostate cancer classification. Huang *et al.* [9] presented a method to identify prostate cancer by the HOG and LBP feature extraction. Farooq *et al.* [10] designed a system using Gabor and LBP with different variations, which showed improved accuracy due to the fusion of different texture features. Although the handcrafted feature methods have achieved great success in this task, the recognition accuracy is limited since handcrafted features can not describe image content in all aspects.

B. Deep Learning based Methods

With the nimble growth of deep learning, prostate cancer detection also benefits from those technologies. In the last decade, many CNN architectures have been conducted by researchers to improve CAD on prostate cancer detection. Deep learning was proved feasible to enhance cancer diagnosis and classification by Fakoor *et al.* [11]. They claimed that deep learning methods have better performance, promising a more comprehensive and generic approach for cancer detection and diagnosis. Then, Wang *et al.* [12] conducted several experiments to demonstrate that deep learning methods generally had better performances than non-deep learning methods by several standards. Shin *et al.* [13] and Tajbakhsh *et al.* [14] proposed pre-trained CNN models. They demonstrated the feasibility of such transfer learning used for medical image processing. Besides 2D images, researches for 3D images are also attempted by researchers. Liu *et al.* [15] developed a novel deep learning architecture called XmasNet was for classification of prostate cancer, using 3D multiple parametric MRI data, and they had satisfactory results. Additionally, some researches focus not on the whole classification task, but on the approaches of one particular phase in the bigger picture. Guo *et al.* [16] decided to go deeper in the segmentation step and they used a sparse patch matching method to infer a prostate likelihood map, which was later used to form the final segmentation.

III. THE PROPOSED METHOD

As is depicted from Fig.2, the model is featured with two separate paths, and each taking charge of deep learning feature extraction and handcrafted feature learning respectively. The model receives data of T2WI, and the images are all pre-processed and reshaped as single-channel grayscale images of size 128×128 . The subnet1 is responsible for deep learning feature extraction and is composed of a ST module (Jaderberg *et al.* [17]) and a ResNet. As for the subnet2, three handcrafted features (HOG, LBP, Haar) are first extracted from the input grayscale images, then a simple CNN to handle all three handcrafted features to select the most effective features are adopted for this particular task. The details of the designing and the combination pattern of these components will be elaborated in subsection III-B.

A. Deep Learning Feature Extraction

Our subnet1 consists of an ST module followed by a ResNet50 with bottleneck structure and it takes the 128×128 grayscale images as input to extract features. A simple ResNet without any modification can set quite a good baseline. Therefore, given the demand of both effectiveness and training speed, we decided to build our model based on a 50 layer ResNet. To further shorten the training time, we utilize the bottleneck residual unit. This bottleneck block has a much lower time complexity, thus saving massive time on training. Additionally, data shortage often happens in medical imaging researches. To solve this problem, we use an ST module which can augment the data in a flexible and learnable way.

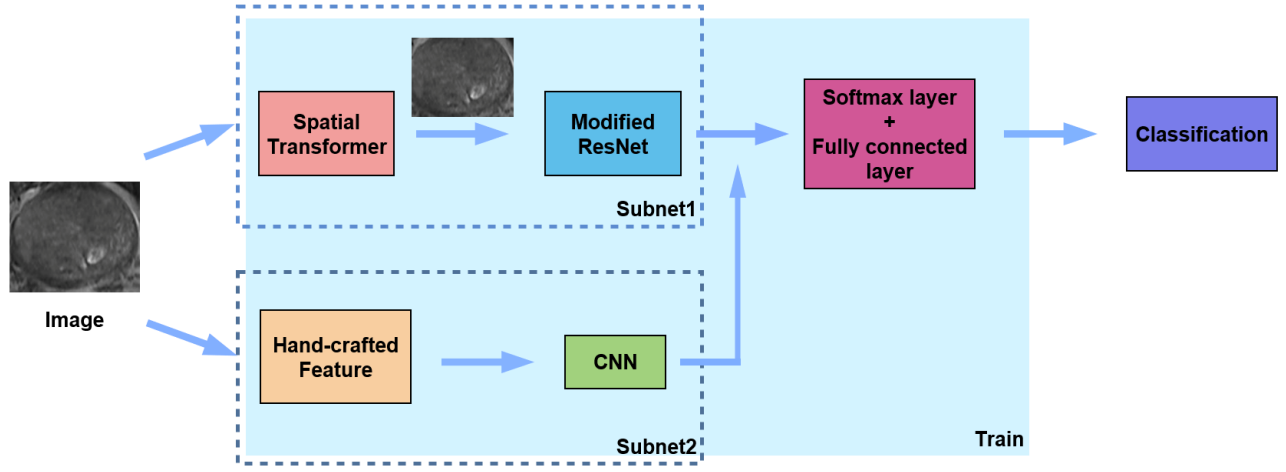


Fig. 2. Main Architecture of Our Model: Subnet1 takes the original image as input to extract high dimension features, while subnet2 first extracts three handcrafted features and then process them with a simple CNN. The outputs of the two subnets are then concatenated and sent to the classifier.

1) *The ST Module*: The ST is a learnable module to do spatial manipulation of data within the network and it can be inserted into any position of existing architecture to boost the network's ability to actively spatially transform the feature map or the original image. This module has three components as in Fig.3: localization network, parameterized sampling grid and differentiable image sampling.

The localization network extracts the features of those input images to find six parameters for affine transformation as in (1). (x, y) is the original coordinate in the input image while (x', y') is the new coordinate in the output feature map. The original affine transformation formula is as in (2). θ is the rotation angle of the input array, and e and f are linear increment parameters. To simplify the issue, the four trigonometric functions can be replaced with four new variables (a, b, c, d) in (1). In the training process, the localization network will optimize the parameters eventually to find the best way to do the affine transformation that is most effective for classification. Also, those parameters and calculation results will be used by the grid generator to acquire the original coordinates. The affine transformation and the grid generator together constitute the parameterized sampling grid. The sampler then uses the coordinates given by the grid generator and the affine transformation to draw the output feature map. To avoid decimals in coordinates, bilinear interpolation is utilized.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (2)$$

Therefore, our main job in constructing our own ST module is to find the suitable localization network and the final sampling size to ensure the best performance of the ST module. The localization net is designed as a CNN with multiple max-pooling, convolution, and activation layers. And several

fully connected layers are placed in the end to output the six parameters. As for the sampling size, the size of 96×96 has been proved most effective after several experiments.

B. Handcrafted Features and Feature Fusion

Although it has been proved that deep learning method outperforms handcrafted feature analysis, handcrafted features can still provide information that can't be covered by CNN extracted features. So it is promising to achieve better performance by introducing handcrafted features to CNN. For this task, we choose HOG and LBP features that provide information on shape and texture respectively for these are the most important basis to detect prostate cancers. Additionally, since the input data is gray-scale images, the Haar feature is employed to acquire the gray level changes of the images.

1) *Handcrafted Feature Extraction*: Commonly used features describe image content from several aspects such as color, shape, and texture. We choose HOG as a shape feature, LBP histogram as a texture feature, and Haar to reflect the gray level changes. Details of the extraction of these features are as followed.

Histogram of Oriented Gradient: HOG feature is a feature descriptor used for object detection in computer vision and image processing. It consists of calculating and statistic the gradient direction histogram of the local area of the image. Compared to other feature description methods, HOG works better in the large spatial domain. The implementation of the HOG feature extraction algorithm first requires pre-processing of the picture, including graying processing and standardization of the color space of the input image using the gamma correction method.

a). From RGB to graylevel image

For a color image, the RGB component is converted into a grayscale image, and the conversion formula is:

$$Gray = 0.3 \times R + 0.59 \times G + 0.11 \times B \quad (3)$$

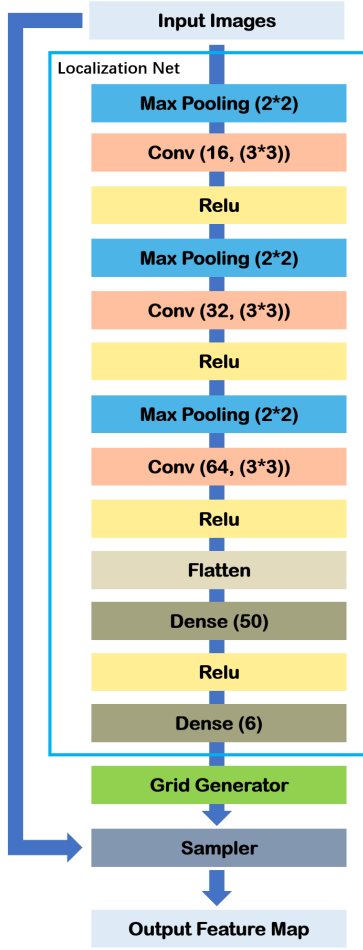


Fig. 3. The architecture of the ST module: The ST module mainly consists of three parts. The localization net decides parameters for affine transformation. Its structure varies according to different tasks. Grid generator uses the affine parameters to locate the original coordinates and sampler uses those coordinates to generate the output feature map.

where RGB is the color representing the three channels of red, green and blue.

b). Gamma correction

In the case of uneven image illuminance, the overall brightness of the image can be increased or decreased by gamma correction. In practice, Gamma standardization can be performed in two different ways, the square root, and the logarithmic method. Here we use the square root method $I(\cdot)$ for gamma correction. The corrected state is Y , and the formula is as follows ($\gamma = 0.5$):

$$Y(m, n) = I(m, n)^\gamma \quad (4)$$

Local Binary Pattern: LBP is an operation used to describe the local texture features of an image; it has significant advantages such as rotation invariance and gray invariance. Moreover, the extracted features are local texture features of the image. The original LBP operator is defined as a window within the window of 3×3 , with the center pixel of the window as a threshold, and the grey value of the adjacent 8 pixels is

compared with it. If the surrounding pixel value is greater than the central pixel value, the pixel point is marked as 1, otherwise 0. In this way, 8 points in the 3×3 neighborhood can be compared to produce an 8-bit binary number (usually converted to a decimal number, ie, LBP code, a total of 256 types), that is, the LBP value of the center pixel of the window is obtained, and this value is used to reflect the texture information for this area. A more formal description of the LBP operator can be given as:

$$LBP(a_c, b_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c) \quad (5)$$

with (a_c, b_c) as central pixel with intensity i_c ; and i_n being the intensity of the neighbor pixel. s is the sign function defined as:

$$s(k) = \begin{cases} 1, & \text{if } k \geq 0 \\ 0, & \text{else} \end{cases} \quad (6)$$

Haar: Haar features are divided into four categories: edge features, linear features, center features, and diagonal features, combined into feature templates. The feature template has two rectangles, white and black, and defines the template's eigenvalue as white rectangle pixels and minus black rectangle pixels. The Haar eigenvalue reflects the gray variation of the image. By changing the size and position of the feature template, a large number of features can be exhaustively listed in the image sub-window. We use the integral map to calculate the Haar feature, which allows us to find the pixel sum of all regions in the image only by traversing the image once, which greatly improves the efficiency of image feature value calculation.

There are two main points in the calculation of the number of Haar features: first, for a certain size feature, sliding calculation in the window; second, for a feature, the feature itself is scaled in the horizontal and vertical directions. The Formula for calculating Haar feature number N :

$$N = XY(W + 1 - w \frac{X+1}{2})(H + 1 - h \frac{Y+1}{2}) \quad (7)$$

Among them, (W, H) is the size of the picture, (w, h) is the size of the rectangle, and $X = \lfloor \frac{W}{w} \rfloor, Y = \lfloor \frac{H}{h} \rfloor$ is the maximum scale factor that the rectangular feature can be enlarged in the horizontal and vertical directions.

After extraction, these three handcrafted features are flattened and then directly connected to one another. After being processed by a simple CNN, the handcrafted features are then concatenated with the features extracted by ST-ResNet to form our final feature map. This feature map is then sent to a fully connected neural network classifier to get the classification result.

IV. EXPERIMENTS

A. Data Introduction

The experiment data is provided by Shanghai Tenth People Hospital, Tongji University. It consists T2WI with fat saturation of 120 patients. The voxel size is $0.9 \times 0.6 \times 3.5 \text{ mm}^3$. 3.0

Tesla (T) whole-body unit MR imaging system (Magnetom Verio 3.0T, Simen Medical Company) was used to acquire those images. The annotations are corroborated by the biopsy pathology. To gain a higher training speed, input images of the proposed model are all resized to 128×128 .

B. Evaluation Metrics

In our experiments, we use accuracy (ACC), sensitivity (TPR), specificity (TNR) and Area Under Curve (AUC) of the receiver operating characteristic curve (ROC) to evaluate the performance of different classification models. Denote TP the true positive, FP the false positive, TN the true negative, and FN the false negative. The definitions of classification accuracy, sensitivity, specificity are as followed:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (8)$$

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$TNR = \frac{TN}{TN + FP} \quad (10)$$

C. Performance with Different Module

As is demonstrated in above, our model can be divided into three modules: the ST module, ResNet50 and the CNN for handcrafted features. Naturally, we shall find out how these modules work and why they function positively with each other.

We compare different models by multiple metrics including accuracy, sensitivity, specificity, and AUC. We start with the most simple one, ResNet50 to set a baseline for this as well as all the experiments to be done. Then the model with the ST module in front of ResNet is evaluated to make sure that the ST module is valid on this task. And finally, we combine these CNN features with selected handcrafted features, which is exactly our proposed model, to find out whether handcrafted features serve as a boost for our model.

The results of this experiment are shown in Table I. Clear improvements on accuracy and AUC of ST-ResNet and our model can be easily depicted. This indicates that the ST module and the handcrafted features we choose are all positively effective. This is because the ST module boosts the model's ability to concentrate on the most effective part of the images, and the handcrafted features can offer information from a much more specific perspective like shape and texture.

TABLE I
PERFORMANCE WITH DIFFERENT MODULE

Size	Accuracy	Sensitivity	Specificity	AUC
ResNet50	0.901	0.918	0.877	0.960
ST-ResNet	0.915	0.918	0.910	0.973
Our Model	0.947	0.964	0.923	0.982

D. The Effect of Different Handcrafted Features

We choose HOG as a shape feature, LBP histogram as texture feature, and Haar based on template feature values as gray-scale changes feature. But it remains to be seen what combination of them can achieve the best performance.

The results of this experiment are shown in Table II. Our model that takes all three handcrafted features has the highest accuracy of 0.947. A single handcrafted feature may not bring obvious enhancement, as we can see that only LBP raises the accuracy from 0.928 to 0.931 while HOG and Haar can't improve the performance much. But when combining these features together, noticeable improvements occur, which is indicated by higher accuracy and more balanced sensitivity and specificity. And the accuracy reaches top when all three handcrafted features are utilized. Therefore, our strategy of selecting and processing handcrafted features is proved effective.

TABLE II
EFFECT OF DIFFERENT HANDCRAFTED FEATURES

Handcrafted Features	Accuracy	Sensitivity	Specificity	AUC
None	0.915	0.918	0.910	0.968
HOG	0.928	0.941	0.910	0.973
LBP	0.931	0.927	0.922	0.979
Haar	0.928	0.955	0.890	0.976
HOG+LBP	0.933	0.932	0.929	0.976
HOG+Haar	0.936	0.955	0.910	0.980
LBP+Haar	0.933	0.955	0.903	0.974
All Three	0.947	0.964	0.923	0.982

E. Comparison with Other CNN Models

To have a more objective evaluation of our model, it is of vital importance to do comparing experiments with other existing well-known CNN models. In this section, we design experiments to show that our model can outperform other state-of-the-art architectures. GoogLeNet [4], ResNet50 [5], InceptionResNetV2 [18], and Xception [19] are chosen as comparing models to make sure that both simple architectures as well as some quite complex models are covered in our experiment.

The results of these experiments are shown in Table III. In terms of Accuracy, there is a climbing trend from the old and simple networks to more complicated ones, which demonstrates that our experiments are designed reasonably. Our model outperforms the others in accuracy, specificity, and AUC. And the sensitivity and specificity are most balanced. So we can safely come to the conclusion that our model outperforms other state-of-the-art CNN architectures in this experiment.

ACKNOWLEDGMENT

This work was supported in part by the General Program of National Natural Science Foundation of China (NSFC) under

TABLE III
COMPARISON WITH OTHER CNN MODELS

Model	Accuracy	Sensitivity	Specificity	AUC
GoogLeNet [4]	0.896	0.955	0.826	0.971
ResNet50 [5]	0.901	0.918	0.877	0.960
InceptionResNetV2 [18]	0.923	0.950	0.884	0.968
Xception [19]	0.928	0.959	0.884	0.968
Our Model	0.947	0.964	0.923	0.982

Grant 61806147, the Shanghai Natural Science Foundation of China under Grant 18ZR1441200, the Fund of the Science and Technology Commission of Shanghai Municipality, China (Grant No. 16411969100) and the State Key Laboratory of Robotics Open Foundation kz0210020192178.

REFERENCES

- [1] C. Prates, S. Sousa, C. Oliveira, and S. Ikram, "Prostate metastatic bone cancer in an egyptian ptolemaic mummy, a proposed radiological diagnosis," *International Journal of Paleopathology*, vol. 1, no. 2, pp. 98–103, 2011.
- [2] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA: a cancer journal for clinicians*, vol. 67, no. 3, pp. 177–193, 2017.
- [3] N. Arumainayagam, H. U. Ahmed, C. M. Moore, A. Freeman, C. Allen, S. A. Sohaib, A. Kirkham, J. van der Meulen, and M. Emberton, "Multiparametric mr imaging for detection of clinically significant prostate cancer: a validation cohort study with transperineal template prostate mapping as the reference standard," *Radiology*, vol. 268, no. 3, pp. 761–769, 2013.
- [4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [6] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [7] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small se-resnet module," *PloS one*, vol. 14, no. 3, 2019.
- [8] G. Nketiah, M. Elschot, E. Kim, J. R. Teruel, T. W. Scheenen, T. F. Bathen, and K. M. Selnæs, "T2-weighted mri-derived textural features reflect prostate cancer aggressiveness: preliminary results," *European radiology*, vol. 27, no. 7, pp. 3050–3059, 2017.
- [9] X. Huang, M. Chen, and P. Liu, "Recognition of transrectal ultrasound prostate image based on hog-lbp," in *2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, pp. 183–187, IEEE, 2019.
- [10] M. T. Farooq, A. Shaikat, U. Akram, O. Waqas, and M. Ahmad, "Automatic gleason grading of prostate cancer using gabor filter and local binary patterns," in *2017 40th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 642–645, IEEE, 2017.
- [11] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proceedings of the international conference on machine learning*, vol. 28, ACM New York, USA, 2013.
- [12] X. Wang, W. Yang, J. Weinreb, J. Han, Q. Li, X. Kong, Y. Yan, Z. Ke, B. Luo, T. Liu, *et al.*, "Searching for prostate cancer by fully automated magnetic resonance imaging classification: deep learning versus non-deep learning," *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.
- [13] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [14] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [15] S. Liu, H. Zheng, Y. Feng, and W. Li, "Prostate cancer diagnosis using deep learning with 3d multiparametric mri," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 1013428, International Society for Optics and Photonics, 2017.
- [16] Y. Guo, Y. Gao, and D. Shen, "Deformable mr prostate segmentation via deep feature learning and sparse patch matching," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1077–1089, 2015.
- [17] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," in *Advances in neural information processing systems*, pp. 2017–2025, 2015.
- [18] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-first AAAI conference on artificial intelligence*, 2017.
- [19] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.