A State-of-the-Art Review on Machine Learning-based Methods for Prostate Cancer Diagnosis

Ari Mohammed ali Ahmed[1], Aree Ali Mohammed[2]

[1]Department of Information Technology, Technical College of Informatics, Sulaimani Polytechnic University, KRG, Sulaimani, Iraq, [2]Department of Computer Science, College of Science, University of Sulaimani, Sulaymaniyah, Iraq

ABSTRACT

Prostate cancer can be viewed as the second most dangerous and diagnosed cancer of men all over the world. In the past decade, machine and deep learning methods play a significant role in improving the accuracy of classification for both binary and multi classifications. This review is aimed at providing a comprehensive survey of the state of the art in the past 5 years from 2015 to 2020, focusing on different datasets and machine learning techniques. Moreover, a comparison between studies and a discussion about the potential future researches is described. First, an investigation about the datasets used by the researchers and the number of samples associated with each patient is performed. Then, the accurate detection of each research study based on various machine learning methods is given. Finally, an evaluation of five techniques based on the receiver operating characteristic curve has been presented to show the accuracy of the best technique according to the area under curve (AUC) value. Conducted results indicate that the inception-v3 classifier has the highest score for AUC, which is 0.91.

Index Terms: Prostate cancer, Machine learning, Deep learning, Algorithm, and Datasets

1. INTRODUCTION

Cancer is a category of diseases that includes cell growth which is irregular with the ability to spread to other areas of the body. Physicists have concentrated on continuous advancement in imaging methods over the past decades, enabling radiologists to improve cancer detection and diagnosis. However, the human diagnosis still suffers from poor repeatability, associated with false identification or perception in clinical decisions of anomalies. Two factors influence these inaccuracies: The ability to observe is limited, for example, perception of human vision is constrained, fatigue duty, or confusion, and the second factor is the clinical case complexity, for instance, unbalanced data which are the mean number of healthy cases are more than a malignant case. Different machine learning-based techniques for cancer detection and classification have introduced a new area of research for early cancer detection. The researches will lead to the ability to reduce the manual system impairments [1]. Another reason, modality that has various analysis techniques such as inappropriate diagnostics, handling, and complicated history is leading to increasing mortality [2].

In the past decades, the field of digital pathology has dramatically developed due to the improvement of algorithms in image processing, machine learning, and advancements in computational power. Within this sector, countless approaches have been suggested to analyze and classify automated pathological images. At present, many

smart and powerful features are added to the microscope and digital images to convert slides of stained tissues into entire digital images. These facilities make a more efficient computer diagnosis system to analyze histopathology and helping early diagnosis. Moreover, they treat cancer by avoiding the increase of cancer cells and easily controlling the tumors from spreading to other parts of the body [3].

In addition, analysis of medical imaging could be significantly involved in identifying defects in various body organs, such as prostate cancer (PCa), blood cancer, skin cancer, breast cancer, brain cancer, and lung cancer. The abnormality of the organ is mainly the result of rapid tumor development, which is the world's leading cause of death. As mentioned by GLOBOCAN statistics, around 18.1 million new cancer cases have appeared in 2018 that gave rise to 9.6 million cancer deaths [2]. PCa is considered the most dangerous disease type of cancer, and it is viewed as the second most commonly diagnosed cancer [3], [4]. The most ubiquitous form of cancer in men is PCa and it has been reported to be the second leading cause of death in men [5].

In the USA, the occurrence of PCa ranks first in men whereas in South Korea, is the fifth most common cancer among males, and the expected cancer deaths in 2018 were 82,155 [3]. PCa is the most leading cancer among men, after lung cancer. It is estimated that about 174,650 new cases and 31,620 PCa-related deaths were recorded in the United States in 2019. PCa considers about 1 in 5 new cancer diagnoses among men. One of the difficulties of PCa is grading that can be considered as a part of the classification problem. Therefore, accurate prediction of PCa grade is crucial to guarantee the quick treatment of malignancy [6]. Furthermore, early diagnosis and treatment planning can significantly reduce the mortality rate due to PCa [6], [7].

Technologies lead to having a crucial role in helping the medical community to diagnose cancer quickly [8]. On the one hand, there are many differences between images attained with modalities of analytic imaging and other image types that related to features and management of procedures. On the other hand, challenges are arising from the use of the different types of scanners, protocols of imaging, variety of noising, and other issues related to image attainment [5].

Different computer-aided techniques have been proposed using a radiomics method or deep learning network to accurately classify the PCa on magnetic resonance imaging (MRI) images [8]. Several studies have shown that computer-aided systems have a remarkable role in PCa detection and

diagnostic evaluation. The methods proposed so far are based on handcrafted features, using a classifier on top to determine whether a PCa lesion is present or to assess its severity by assigning a specific class label. Recently, different techniques such as convolutional neural networks (CNN), support vector machine (SVM), iterative random (random forest [RF]), and J48 in the field of machine learning are proposed for locating and identifying cancer cells and normal cells. They have shown an impressive performance in various computer vision tasks following training with large image databases [5], [9].

This paper aims to propose a state-of-the-art review that surveys several techniques for PCa diagnosis, moreover, the techniques which are mostly based on machine learning are comparing in terms of performance accuracy.

The structure of the paper is as follows: In Section II, a review of some related works is represented while in Section III, the methodology of the literature review is described. Section IV shows a comparison among the aforementioned methods. Finally, a conclusion and future direction of the research survey are given in Section V.

2. SURVEY OF PCA TECHNIQUES

Several techniques have been suggested by many researchers for improving and developing PCa detection. In this survey, we mainly focus on the researcher's techniques that have been implemented with the machine learning field between 2015 and 2020.

Sammouda *et al.*, 2015, worked on malignant PCa cells using near-infrared optical imaging technique that uses the high absorption of hemoglobin in PCa cells. Two algorithms (k-mean and fuzzy clustering mean) are used to segment and extract the cancer region in the prostate's infrared images. Using the Student's t-test to measure the accuracy between these two clusters, P value of K-means "cluster 3" is < 0.0001, and the standard error = 0.0002 is less than P value of ferric carboxymaltose (FCM)<0.0252 and standard error = 0.004. As the result, the K-mean is more accurate than FCM based on statistical analysis [10].

Mohapatra and Chakravarty, 2015, suggested a model using three classifiers SVM, Naive Bayes, and KNN to classify PCa. In this model, microarray is used as a dataset. The area under the curve (AUC) and accuracy have been measured to compare and evaluate the performances of these classifiers,

with taken the entire datasets and selected optimal features separately as the input to the classifiers one by one. As the result, the SVM technique performs more efficacy with higher accuracy of 95.5% [11].

In the same year, Bouazza *et al.*, 2015, proposed a classification method that performed a comparative study of four feature selection methods Fisher, T-Statistics, SNR, and ReliefF, using two classifiers K-nearest neighbors and SVM. Test results indicated that the best classification accuracy is obtained with SVM classifier and SNR method [12].

Dash *et al.*, 2016, worked on the microarray medical datasets and, two variations of kernel ridge regression (KRR) are used which are WKRR and RKRR to classify the datasets. To achieve a high rate of accuracy, this model is comparing the accuracy test among several techniques such as KRR, SVM-RBF, SVM-POLY, and RF. As the result, KRR (WKRR and RKRR) has been higher than all of them, especially RKRR which has an accuracy rate of 97%. However, this model has some drawbacks related to feature extraction which are ignoring the interaction with the classifier, features are considered independently which is mean ignoring these features which are dependencies. Another difficulty is related to determine the point of threshold to rank the features [13].

Imani *et al.*, 2016, proposed an approach to integrate mp-MRI with temporal ultrasound for PCa classification, in vivo. CNN technique has been utilized in this approach. A combination of mp-MRI and temporal ultrasound is used to reduce the missing regions of tumors. The AUC of 0.89 has been achieved for the classification of cancer with higher grades. Despite the importance of this model, there are some drawbacks because of the heterogeneous of PCa and it is difficult to determine tissue signature consistently [14].

Ram *et al.*, 2017, proposed an iterative RF (iRF) algorithm as a classifier model to separate cancer from the controlled samples of PCa. The method worked on microarray and next-generation sequencing (NGS) data. However, having a large number of gene expression data make it difficult of how to identify the biomarkers related to cancer. The RF has been used to select the genes which can diagnose and treat cancer effectively. RF method is used to extract very small sets of genes while it is taken predictive performance. Genes of *SNRPA1* are selected for PCa with the obtained accuracy of 73.33% [15].

Sun *et al.*, 2017, suggested a model investigate the performance of SVM algorithm and to predict the prostate tumor location

using multiparametric MRI data. The capability of best predictive is achieved by optimizing model parameters using leave-one-out cross-validation. A binary SVM classifier utilizes to find a plane in feature space, frequently identified as a decision boundary, which splits the data into two parts. Furthermore, this algorithm is used to search for a decision boundary that maximizes the margin between the two groups. The final model gives results of classification by predicting the higher accuracy of 80.5%. However, only signal intensities and values from both T2-weighted (T2w) images and parametric maps are incorporated as features, respectively [16].

Liu and An, 2017, suggested a model based on deep learning and CNN for image classification of PCa, they used diffusion-weighted MRI (DWI) images that are selected images from a number of patients including positive and negative images. However, a small dataset makes a difficult for training a model that achieves higher accuracy. The proposed model has yielded an accuracy of 78.15% [17].

Reda *et al.*, 2018, presented a model using CNN based on computer-aided design (CAD) system for early diagnosis of PCa from DWI. They achieved accuracy rate of 95.65% [18].

Bhattacharjee *et al.*, 2019, developed a system for digitized histopathology images using a supervised learning method. SVM has been presented and used to classify malignant and benign PCa Grade 3, achieved accuracy was 88.7%. In SVM classification, 2-fold cross-validation has been used to train the model. Both of linear and Gaussian kernel are used for classifying samples as benign and malignant. Furthermore, a binary classification approach has been used which divides the multitype classification into two-category groups. Each partition characterizes distinct and independent classifications which are malignant and benign [3].

Yoo *et al.*, 2019, proposed a model of CAD system based on CNN and RF techniques for MRI (DWI) images. Five individually trained CNNs have been used to categorized DWI slices to extract the features and RF classifier has been used to classify patients into two groups patient with PCa and without PCa with achieved 0.84 as an AUC. The main limitation of this model is intrinsically biased which is mean these patients take MRIs who have symptoms of PCa. On the other hand, by depending on the reports of radiology, these slices with no biopsy consider as a negative sample while slices with biopsy consider as a positive sample based on pathology reports [19].

Cahyaningrum *et al.*, 2020, proposed a method of artificial neural network (ANN) that optimized by genetic algorithm (GA) for PCa detection. This approach gives 76.4% of accurate detection. ANN has some limitations because of involving a huge number of parameters. Consequently, there have been many efforts to fix some of these limitations by joining ANN with another algorithm to address this problem. GA is an algorithm that is compatible and adapted with the ANN algorithm [20].

Besides, Duran-Lopez *et al.*, 2020, presented a novel CAD system based on a deep learning algorithm (CNN) for distinguishing between malignant and normal tumors in whole slide images (WSIs). Cross-validation technique has been used with patches extracted from WS images. In this approach, the higher accuracy rate has achieved 99.98% [21].

Liu *et al.*, 2020, stated a model of deep learning that integrates S-Mask R-CNN with Inception-v3 in ultrasound images to diagnose PCa. Furthermore, the AUC for Inception-v3 is 0.91. According to this model, there is a lot of traditional classifiers that can be used such as SVM and K-nearest neighbor. Due to minor variation between the ultrasonic PCa images and serious noise interface, some miss classifications might happen. Therefore, the CNN was presented in deep learning to achieve the best improvement of the classification accuracy in ultrasound images of the prostate without needing to describe the features manually and target image extraction [22].

3. METHODOLOGY

This review paper has conducted various studies in the field of PCa that is based on machine and deep learning techniques. First, the datasets that have been used by the researchers are described in subsection A. While in subsection B, the methodologies that are related to the performance accuracy are explained.

3.1. Datasets
Different datasets that have been used by researchers were investigated. Table 1 shows the modality of the dataset types that have been used by the authors in this survey. Moreover, sample numbers associated with each patient were given.

3.2. Performance parameter criteria
In the classification process, performance measurement is very important and essential, which determines the accuracy of the model. For this purpose, receiver operating

TABLE 1: Datasets types with number of samples

Authors	Modality	No. of samples
Mohapatra and Chakravarty[11]	Microarray	136
Dash *et al.* [13]	Microarray	136
Ram *et al.* [15]	Microarray Gene Expression Omnibus (GSE71783)	30
Sun *et al.* [16]	MRI (T2w, DWI and DCE)	5
Liu and An [17]	MRI (DWI)	200
Reda *et al.* [18]	MRI (DWI)	23
Bhattacharjee *et al.* [3]	Microscopic tissue images	400
Cahyaningrum *et al.* [20]	Microarray gene expression	102
Duran-Lopez *et al.*[21]	WSI (whole slide image)	97

characteristic (ROC) and AUC are proposed as effective evaluation metrics based classification model's performance.

In statistics, a ROC curve can be defined as a graphical plot to illustrate the performance of a binary classification as it is used to distinguish varied thresholds. The true-positive rate (TPR) against the false-positive rate (FPR) is plotted to create the curve at varied threshold settings. In machine learning, the terms of recall, sensitivity, or detection probability have the same meaning as TPR. While, the term of fallout or false alarm probability has the same meaning as FPR and could be calculated as (1-specificity) [23]. Fig. 1 illustrates the relation between the ROC and AUC.

Furthermore, ROC is the probability curve whereas AUC is the degree of separable classes. ROC indicates that how much the model is capable of distinguishing amongst classes. Higher the AUC value (between 0 and 1) leads to better accuracy of the model.

This survey compares the techniques that are based on the accuracy of the proposed methods. Confusion matrix (CM) with performance metrics such as specificity and sensitivity is used to evaluate the proposed models [24]. The CM output could be either binary or multiclass. It has also a table of four different combinations between actual and predicted values. Predicted values are predicted by the model while actual values are actually in a dataset. Fig. 2 shows the CM relations.

The following formulas describe the performance accuracy metrics based on TP, TN, FP, and FN, according to CM.
TP – Values that are actually positive and predicted positive.
FP – Values that are actually negative but predicted to positive.

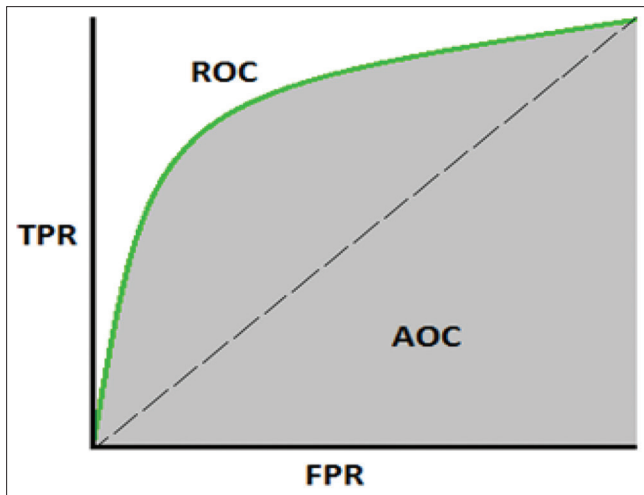FN – Values that are actually positive but predicted to negative.

Fig. 1. Area under the curve-receiver operating characteristic curve.

Fig. 2. Confusion matrix combinations.
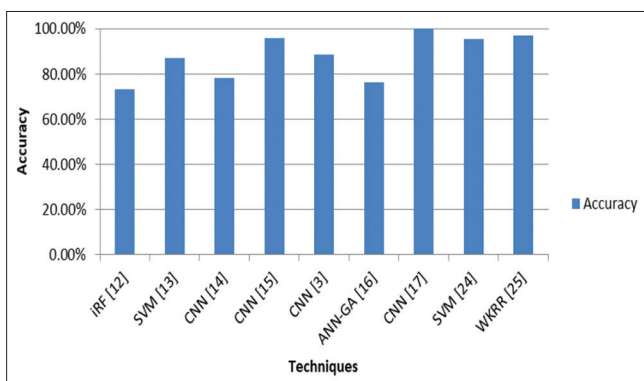
Fig. 3. Accuracy comparison of each technique.

TN – Values that are actually negative and predicted to negative.

$$\text{TPR} \left(\text{Sensitivity or recall}\right) = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (2)$$

$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{FP + TN} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

TP and TN represent the number of correctly predicted positive and negative samples, while FP and FN are used to represent the number of incorrectly predicted positive and negative samples [25].

4. COMPARISON AND DISCUSSION

Many different techniques have been used by researchers. Each technique used a special type of dataset. Here, we compare the methods based on the accuracy with the dataset types and the year of publication, as shown in Table 2.

Fig. 3 shows the accuracy of the techniques separately.

Finally, an evaluation of five techniques based on the AUC has been performed to show the accuracy of the best technique, as depicted in Fig. 4.

As a result, according to the AUC measurements, the Inception-v3 classifier has the highest score for AUC, which

TABLE 2: Techniques with accuracy

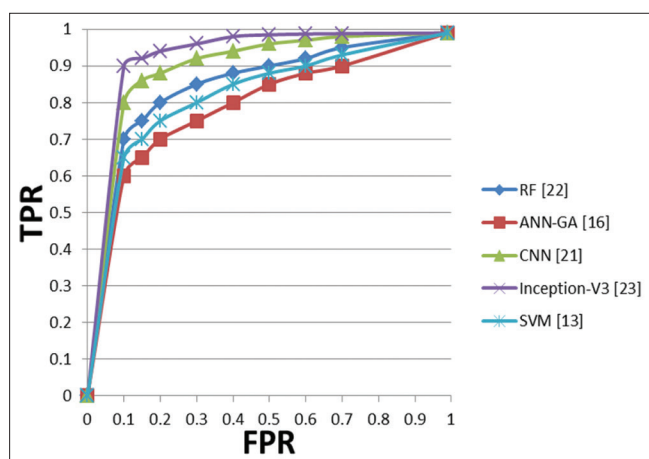| S. No.	References	Year	Methods	Accuracy percent
1	Mohapatra and Chakravarty[11]	2015	SVM	95.5
2	Dash et al. [13]	2016	WKRR	97
3	Ram et al. [15]	2017	iRF	73.3
4	Sun et al. [16]	2017	SVM	80.5
5	Liu and An [17]	2017	CNN	78.15
6	Reda et al. [18]	2018	CNN	95.65
7	Bhattacharjee et al. [3]	2019	CNN	88.7
8	Cahyaningrum et al. [20]	2020	ANN-GA	76.4
9	Duran-Lopez et al. [21]	2020	CNN	99.98

Fig. 4. Receiver operating characteristic curve for five classifiers.

is 0.91, although the type and quality of the dataset affect the ratio of the AUC scale.

5. CONCLUSION

This paper has introduced a comparison of classification methods based on machine learning techniques of the research related to PCa using various datasets including (Microarray, Microarray Gene Expression Omnibus (GSE71783), MRI (T2w, DWI, and DCE), microscopic tissue images, and WSI. In addition, the methods used in the literature have been reviewed along with the available results of the performance accuracy. The higher value of the AUC is identified amongst most five recent papers and it is 0.91.

REFERENCES

[1] Lemaître, R. Martí, J. Freixenet, J. C. Vilanova, P. M. Walker and F. Meriaudeau. "Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review". *Computers in Biology and Medicine*, vol. 60, pp. 8-31, 2015.

[2] T. Saba. "Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges". *Journal of Infection and Public Health*, vol. 13, no. 9, pp. 1274-1289, 2020.

[3] S. Bhattacharjee, H. G. Park, C. H. Kim, D. Prakash, N. Madusanka, J. H. So, N. H. Cho and H. K. Choi. "Quantitative analysis of benign and malignant tumors in histopathology: Predicting prostate cancer grading using SVM". *Applied Sciences*, vol. 9, no. 15, 2019.

[4] S. Liu, H. Zheng, Y. Feng and W. Li. "Prostate cancer diagnosis using deep learning with 3d multiparametric MRI". *SPIE Proceedings*, vol. 10134, pp. 3-6, 2017.

[5] N. Aldoj, S. Lukas, M. Dewey and T. Penzkofer. "Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network". *European Radiology*, vol. 30, no. 2, pp. 1243-1253, 2020.

[6] B. Abraham and M. S. Nair. "Automated grading of prostate cancer using convolutional neural network and ordinal class classifier". *Informatics in Medicine Unlocked*, vol. 17, p. 100256, 2019.

[7] L. A. Torre, B. Trabert, C. E. DeSantis, K. D. Miller, G. Samimi, C. D. Runowicz, M. M. Gaudet, A. Jemal, R. L. Siegel. "Ovarian cancer statistics, 2018". *CA: A Cancer Journal for Clinicians*, vol. 68, no. 4, pp. 284-296, 2018.

[8] M. Arif, I. G. Schoots, J. C. Tovar, C. H. Bangma, G. P. Krestin, M. J. Roobol, W. Niessen and J. F. Veenland. "Clinically significant prostate cancer detection and segmentation on low-risk patients using a convolutional neural network on multi-parametric MRI". *European Radiology*, vol. 30, pp. 6582-6592, 2020.

[9] L. Brunese, F. Mercaldo, A. Reginelli and A. Santone. "Formal methods for prostate cancer Gleason score and treatment prediction using radiomic biomarkers". *Magnetic Resonance Imaging*, vol. 66, pp. 165-175, 2020.

[10] R. Sammouda, H. Aboalsamh and F. Saeed. "*Comparison Between K Mean and fuzzy C-mean Methods for Segmentation of Near Infrared Fluorescent Image for Diagnosing Prostate Cancer*". International Conference on Computer Vision and Image Analysis Applications, 2015.

[11] P. Mohapatra and S. Chakravarty. "*Modified PSO Based Feature Selection for Microarray Data Classification*". 2015 IEEE Power, Communication and Information Technology Conference, pp. 703-709, 2015.

[12] S. H. Bouazza, N. Hamdi, A. Zeroual and K. Auhmani. "*Gene-expression-based Cancer Classification through Feature Selection with KNN and SVM Classifiers*". 2015 Intelligent Systems and Computer Vision, 2015.

[13] P. Mohapatra, S. Chakravarty and P. K. Dash. "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system". *Swarm and Evolutionary Computation*, vol. 28, pp. 144-160, 2016.

[14] F. Imani, S. Ghavidel, P. Abolmaesumi, S. Khallaghi, E. Gibson, A. Khojaste, M. Gaed, M. Moussa, J. A. Gomez, C. Romagnoli, D. W. Cool, M. Bastian-Jordan, Z. Kassam, D. R. Siemens, M. Leveridge, S. Chang, A. Fenster, A. D. Ward and P. Mousavi. "*Fusion of Multi-parametric MRI and Temporal Ultrasound for Characterization of Prostate Cancer: In vivo Feasibility Study*". Medical Imaging 2016: Computer-Aided Diagnosis, vol. 9785, p. 97851K, 2016.

[15] M. Ram, A. Najafi and M. T. Shakeri. "Classification and biomarker genes selection for cancer gene expression data using random forest". *The Iranian Journal of Pathology*, vol. 12, no. 4, pp. 339-347, 2017.

[16] Y. Sun, H. Reynolds, D. Wraith, S. Williams, M. E. Finnegan, C. Mitchell, D. Murphy, M. A. Ebert and A. Haworth. "Predicting prostate tumour location from multiparametric MRI using Gaussian kernel support vector machines: A preliminary study". *Physical and Engineering Sciences in Medicine*, vol. 40, no. 1, pp. 39-49, 2017.

[17] Y. Liu and X. An. "*A Classification Model for the Prostate Cancer Based on Deep Learning*," Proceedings of the 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017, pp. 1-6, 2018.

[18] I. Reda, A. Shalaby, M. Elmogy, A. A. Elfotouh, F. Kahalifa, M. A. El-Ghar, E. Hosseini-Asl, G. Gimel'farb, N. Werghi and A. El-Baz. "*A New CNN-Based System for Early Diagnosis of Prostate Cancer*". Proceedings International Symposium on Biomedical Imaging, pp. 207-210, 2018.

[19] S. Yoo, I. Gujrathi, M. A. Haider and F. Khalvati. "Prostate cancer

detection using deep convolutional neural networks". *Scientific Reports*, vol. 9, no. 1, pp. 1-10, 2019.

[20] K. Cahyaningrum, Adiwijaya and W. Astuti. "*Microarray Gene Expression Classification for Cancer Detection using Artificial Neural Networks and Genetic Algorithm Hybrid Intelligence*," 2020 International Conference on Data Science and its Applications, 2020.

[21] L. Duran-Lopez, J. P. Dominguez-Morales, A. F. Conde-Martin, S. Vicente-Diaz and A. Linares-Barranco. "PROMETEO: A CNN-based computer-aided diagnosis system for WSI prostate cancer detection". *IEEE Access*, vol. 8, pp. 128613-128628, 2020.

[22] . Liu, C. Yang, J. Huang, S. Liu, Y. Zhuo and X. Lu. "Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer". *Future Generation Computer Systems*, vol. 114, pp. 358-367, 2021.

[23] A. Z. Shirazi, S. J. S. Mahdavi Chabok and Z. Mohammadi. "A novel and reliable computational intelligence system for breast cancer detection". Medical and Biological Engineering and Computing, vol. 56, no. 5, pp. 721-732, 2018.

[24] M. Nour, Z. Cömert and K. Polat. "A novel medical diagnosis model for COVID-19 infection detection based on deep features and bayesian optimization". *Applied Soft Computing*, vol. 97, pp. 1-13, 2020.

[25] Y. Celik, M. Talo, O. Yildirim, M. Karabatak and U. R. Acharya. "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images". *Pattern Recognition Letters*, vol. 133, pp. 232-239, 2020.