Prostate Cancer Grading From Histopathology Images Using MobileNet

03CS7914 Project (Phase II)

CHN20MT019 CHN20CSIP07 Sree Lekshmi B S
sreelekshmibs16@gmail.com
M. Tech. Computer Science & Engineering (Image Processing)

Department of Computer Engineering
College of Engineering Chengannur
Alappuzha 689121
Phone: +91.479.2165706
http://www.ceconline.edu
hod.cse@ceconline.edu

College of Engineering Chengannur
Department of Computer Engineering

C E R T I F I C A T E

This is to certify that, this report titled ***Prostate Cancer Grading From Histopathology Images Using MobileNet*** is a bonafide record of the work done by

CHN20MT019 CHN20CSIP07 Sree Lekshmi B S

Fourth Semester M. Tech. Computer Science & Engineering (Image Processing)

student, for the course work in **03CS7914 Project (Phase II)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, M. Tech. Computer Science & Engineering (Image Processing) of **APJ Abdul Kalam Technological University**.

Guide

Gopakumar G
Associate Professor
Computer Engineering

Coordinator

Ahammed Siraj K K
Associate Professor
Computer Engineering

Head of the Department

July 18, 2022

Dr. Manju S Nair
Associate Professor
Computer Engineering

Permission to Use

In presenting this project dissertation at College of Engineering Chengannur(CEC) in partial fulfillment of the requirements for a postgraduate degree from APJ Abdul Kalam Technological University, I agree that the libraries of CEC may make it freely available for inspection through any form of media. I further agree that permission for copying of this dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the Head of the Department of Computer Engineering. It is understood that any copying or publication or use of this dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to CEC in any scholarly use which may be made of any material in this project dissertation.

Sree Lekshmi B S

Statement of Authenticity

I hereby declare that this submission is my own work except the dataset used in the work and to the best of my knowledge it contains no other materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at College of Engineering Chengannur(CEC) or any other educational institution, except where due acknowledgement is made in the report. Any contribution made to my work by others, with whom I have worked at CEC or elsewhere, is explicitly acknowledged in the report. I also declare that the intellectual content of this report is the product of my own work done as per the **Problem Statement** and **Proposed Solution** sections of the project dissertation report. I have explicitly stated the major references of my work. I have also listed all the documents referred, to the best of my knowledge.

Sree Lekshmi B S

Acknowledgements

Primarily, I thank Lord Almighty for his eternal support through out my project work.

I would like to convey my heartfelt thanks to Dr.Smitha Dharan, Principal, College of Engineering Chengannur for giving all the facilities required for doing my project work. My heartfelt words of gratitude to Dr. Manju S Nair, Professor and Head of Department of Computer Engineering, for providing constant support.

I express my heartfelt gratitude to my Project Co-ordinator Mr.Ahammed Siraj K K, Associate Professor in Computer Engineering, Mr.Gopakumar G, Associate Professor in Computer Engineering, for their timely suggestions and expert guidance.

Now I would like to express my grateful thanks to Dr. K Sujathan, Associate Professor, research expertise in Tumor Biology-Proteomics, Cytopathology at RCC, who helped me lot in gathering different information, collecting data and guiding me from time to time. Also thanking him to giving me histopathology image dataset of prostate cancer. Email: ksujathan@gmail.com

Sree Lekshmi B S

Abstract

Image segmentation and classification have significant relevance in the field of medical image analysis for grading and diagnosing diseases. Histopathology describes the process by which pathologists examine biopsy samples under a microscope to locate, analyze and classify the majority of diseases, such as cancer. Pathologist's subjective assessment of histopathology image analysis depends on examiner level of knowledge. Pathologists look at the cell distribution and tissue patterns to identify benign and malignant lesions in images as well as to forecast the stage of disease. This takes a lot of time and the prediction might not be accurate. For a quantitative diagnosis of tissue, a computer assisted histopathological image analysis is required to get around this. Prostate Cancer is one of the worst diseases among males. The proposed work segments and classifies histopathology images to find the tissue patterns for prostate cancer diagnosis in three steps:1)pre-processing, 2)gland segmentation and 3) classification(grading). Images are pre-processed with principal component analysis followed by k-means clustering method. The method employs image classification with a deep learning network for classifying prostate cancer into three grades ie grade 3, 4 and 5. The performance evaluation of the proposed method is performed and obtains an accuracy of 93%.

Contents

List of Figures

Chapter 1

Introduction

Histopathology is a scientific term which is derived from histology and pathology. Histology deals with the study of microscopic structures of cells or tissues of organisms and pathology is the study of diseases. So histopathology refers to the microscopic examination of biopsy or a surgical specimen obtained from a patient body that is properly dyed with stains in glass slides, under a microscope with specialized camera to locate and study signs of diseases such as cancer. Understanding microscopic structures and their roles at the cellular, sub cellular, tissue and organ level is crucial to understanding the progression and prognosis of diseases. A pathologist can determine the morphological properties of tissue structures by examining histopathological samples under a microscope at different magnification levels. This allows them to determine how abnormal your cells appear and how likely it is that the disease will progress and spread. When a disease is identified, a grading procedure is used to establish how widely it has spread. A grading strategy is planned in light of this. Analysis of histopathology images is typically done to find and identify diseases like cancer. Since the diagnosis in pathology reports is dependent on the pathologists' subjective assessments, it may vary depending on the examiner's level of knowledge. It takes a lot of time and is less accurate prediction. So employing computer assisted systems will solve this issue. Computer assisted systems can make quantitative assessments of images which is useful for objective diagnosis.

After obtaining a digital histology image from a biopsy sample, a diagnosis can vary depending on how the images are manually examined. Computer-assisted systems that provide unbiased analyses of diseases are used to solve this problem. To build a computer assisted analytical system, there are some fundamental process are needed to be implemented on the system. They are pre-processing and segmentation methods, feature extraction and selection methods and disease detection, classification and post-processing methods.

The histopathological image analysis includes the computations performed at different magnifications ($2\times$, $4.5\times$, $10\times$, $20\times$ and $40\times$) for multivariate statistical analysis, diagnosis and classification. It can be done at lower magnification for tissue level analysis. Using image pre-processing, feature extraction, and classification approaches such thresholding, morphological processing, region and boundary-based analysis and supervised classification techniques, pathologists examine histopathology pictures. The creation of quantitative and automated computerised image analysis algorithms is required by the current advancement of digital pathology in order to help pathologists evaluate

the numerous digitised histopathology images. Depending on the application or kind of disease the image processing steps may vary, but in general the image processing algorithms are similar for most of the applications. Pathologist study histology reports of various organs in human body like liver, colon, breast, prostate etc to detect presence of cancers. Prostate cancer is one of the serious disease facing men in the world.

In the male reproductive system, the prostate is a little walnut-shaped gland that sits below the bladder and surrounds the urethra. It makes the seminal fluid, which feeds and carries sperm. Non-glandular stroma and glands that are located around the stroma make up healthy prostate tissue. A joint capsule encloses and securely fuses these various tissues. Lumen and epithelial cells make up each gland unit. Prostate cancer most frequently develops in the outside, peripheral region of the gland. In malignant tissues, cells grow in and out of the gland, disrupting the prostate gland's normal structure and organisation. Epithelial cells reproduce uncontrollably in cancerous tissue, disrupting the normal positioning of gland units. Prostate cancer depends on age, family history

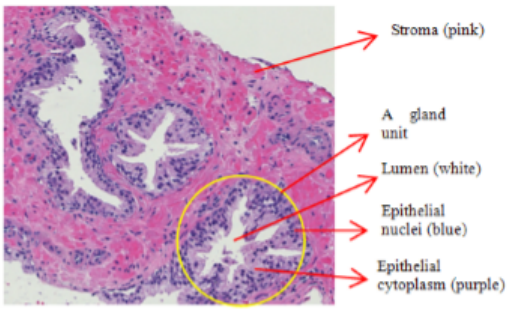

Figure 1.1: Histopathology Image of Normal Prostate

and life style. This can be occured in men after the age of 50. Symptoms of PCa includes difficulty in urination, blood in urine and pain in pelvis. Main cause to prostate cancer is due to gene changes, men who eat lot of red meat, fat products etc. Statistical data of prostate cancer shows that a huge number of people were diagnosed prostate cancer every year and also the death rate is high. So detection, prognosis prediction and treatment planning for this cancer is essential. There are several methods to diagnose prostate cancer by using screening tests for urinary symptoms.

The digital rectal exam (DRE), which was first used to detect prostate cancer in the early 1990s. By sticking a finger into the patient's anus, the doctor determines whether the patient's prostate is enlarged using this method. PSA is a screening test used for prostate cancer. Prostate cancer screening tests use PSA (prostate-specific antigen). The PSA test can help some patients receive the necessary therapy prior to the cancer spreading by identifying early-stage prostate cancer, especially in those who have many risk factors. False-positive results from PSA testing are common, along with health issues like discomfort and infection. Another choice is a biomarker test. A biomarker is a molecule that has been found in the blood, urine or body tissues of a cancer patient. It is created by the tumour or the immune system of the body in response to cancer. Two biomarker tests for prostate cancer include the 4K score, which forecasts the probability of high-risk prostate

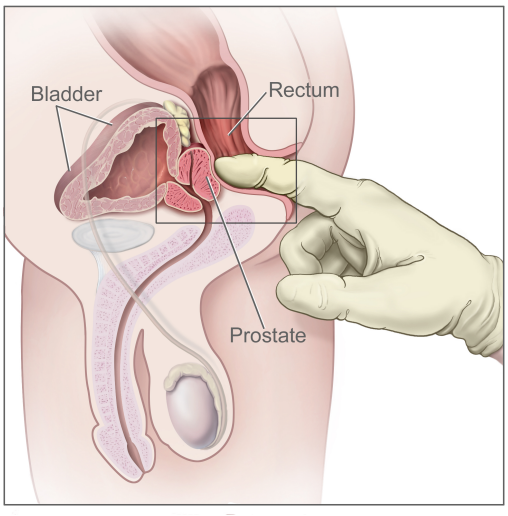

Figure 1.2: Digital Rectal Exam

cancer and the Prostate Health Index (PHI), which forecasts the probability of prostate cancer. Additional testing is required to identify prostate cancer if the results of the PSA and DRE tests are abnormal. Therefore, biopsy is frequently recommended. A small amount of prostate tissue is taken out and placed on a glass slide. The glass slide is then appropriately dyed and put under a microscope to be studied. This technique is usually performed in a hospital setting. Before the biopsy, the patient is given local anaesthetic. To collect tissue samples, the ultrasonography probe is put into the rectum and then the biopsy needle is pushed through the rectum and into the prostate gland.

1.1 Gleason Grading

Gleason grading is a well-known method pathologists use to diagnose prostate cancer. In 2004, World Health Organization(WHO) approved the Gleason grading system. The Gleason grading system is used to help evaluate the prognosis of men with prostate cancer using samples from a prostate biopsy. Together with other parameters, it is incorporated into a strategy of prostate cancer staging which predicts prognosis and helps guide therapy. A Gleason score is given to prostate cancer based upon its microscopic appearance. Cancers with a higher Gleason score are more aggressive and have a worse prognosis. Pathological scores range from 2 to 10, with higher numbers indicating greater risks and higher mortality.

Gleason grading system consists of five histological growth patterns. Prostate cancer is divided into five grades according to the Gleason grading system. The histology of grade 1 is characterized by a dense arrangement of well-differentiated glands, forming well-defined nodules. Grade 2 shows a more loose arrangement of well differentiated glands forming clearer nodules. Grade 3 is characterized by scattered, independent moderately differentiated glands. Grade 4 manifests as poorly differentiated, fused glands. Grade 5 is characterized by a lack of glandular differentiation. The high grade number means the prognosis is worse. The results show that the deep learning approach has great potential for grading prostate cancer from histopathology images.

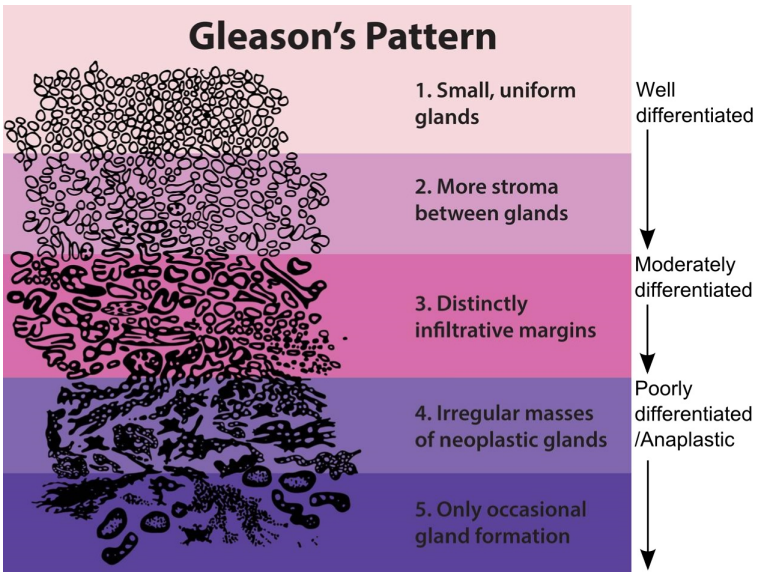

Figure 1.3: Gleason Patterns

This paper proposes an effective method using deep learning approach for classification and segmentation of histopathology images. The project focus on detecting and identify the severity of prostate cancer from histopathology images using gleason grading method. For detection first enhance the image using median filtering. This work considers domain knowedge about gland structures, the nuclei and stroma to improve the sensitivity and selectivity of the classification. As the shape and position of the three main parts of a gland (nuclei, cytoplasm and lumen) are diverse, first segment the three gland components based on shape, size and color. Then unify them to segment the full gland. After segmenting the glandular structures, features are extracted and gleason grading is performed to detect and determine the severity of prostate cancer using deep learning network.

1.2 Proposed Project

This project proposes a deep learning approach for prostate cancer grading into three main grades from histopathology images using Gleason grading method. In this work morphological features are well analyzed for determining how advanced your prostate cancer is. The project work consists of three steps, 1)Pre-processing, to enhance images using median filter, 2)Gland segmentation, to segment various components of prostate tissues and 3)Classification, to grade prostate prostate cancer into grade 3, 4 or 5.

1.2.1 Problem Statement

Grading of prostate cancer using deep learning approach by examining morphological features of gland units.

1.2.2 Proposed Solution

Segment glandular structures using pixel classification and boundary extraction method, and then grading of prostate cancer using MobileNet.

Chapter 2

Report of Preparatory Work

2.1 Literature Survey Report

1. **A new approach to diagnosing prostate cancer through magnetic resonance imaging**, Elsevier, Oct. 2020, pp. 897-904.

   In this paper a combination of improved GrowCut, Zernik feature extraction and ensemble learning techniques such as KNN, SVM and MLP algorithms based on the geometric properties of MRI images were used for prostate cancer detection and lesion segmentation. Using the improved growcut method they seperate the image of prostate cancer from MRI based on geometrical features. With the help of Zernik technique, image features were extracted based on geometrical, edge, and texture properties of the images and finally, by combining machine learning methods including KNN, SVM and MLP, the lesion was determined and diagnosed in prostate cancer MRI images. After simulating the proposed method, they found that the accuracy of this method, which combines several methods, improved by about 20% compared to other methods.

2. **Microarray Core Detection by Geometric Restoration**, Analytical Cellular Pathology, vol 35, pp. 381-393, 2012.

   This paper explains about Whole-slide imaging of tissue microarrays (TMAs) which holds the promise of automated image analysis of a large number of histopathological samples from a single slide. This demands high-throughput image processing to enable analysis of these tissue samples for diagnosis of cancer and other conditions. In this paper, they present a completely automated method for the accurate detection and localization of tissue cores that is based on geometric restoration of the core shapes without placing any assumptions on grid geometry. The method relies on hierarchical clustering in conjunction with the Davies-Bouldin index for cluster validation in order to estimate the number of cores in the image where from we estimate the core radius and refine this estimate using morphological granulometry. The final stage of the algorithm reconstructs circular discs from core sections such that these discs cover the entire region of each core regardless of the precise shape of the core. The results show that this method is able to reconstruct core locations without any evidence of localization.

3. **Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer** ,, IEEE Access, July. 2020, pp. 117714–117725.

   This paper presents an automated Gleason grading and Gleason pattern region segmentation method based on deep learning for pathological images of prostate cancer. An architecture combining the atrous spatial pyramid pooling and the multi scale standard convolution is proposed for the segmentation of the Gleason pattern region to get accurate Gleason grading. In addition, the post processing procedure based on conditional random fields is applied to the prediction. The quantitative experiments on 1211 prostate cancer tissue microarrays demonstrate that their results have a high correlation with the manual segmentations. The mean intersection over union and the overall pixel accuracy for the Gleason pattern region are 77.29% and 89.51%, respectively. Furthermore, the results of the automatic Gleason grading were comparable to the results of experienced pathologists.

4. **Histological stain evaluation for machine learning applications**, International conference on Medical Image Computing and Computer Assisted Intervention, Oct. 2012.

   This work is a methodology for quantitative comparison of histological stains based on their classification and clustering performance, which may facilitate the choice of histological stains for automatic pattern and image analysis. Machine learning and image analysis are becoming increasingly important in pathology applications for automatic analysis of histological tissue samples. Pathologists rely on multiple, contrasting stains to analyze tissue samples, but histological stains are developed for visual analysis and are not always ideal for automatic analysis. In this paper thirteen different histological stains were used to stain adjacent prostate tissue sections from radical prostatectomies. They evaluate the stains for both supervised and unsupervised classification of stain/tissue combinations. For supervised classification they measure the error rate of nonlinear support vector machines, and for unsupervised classification use the Rand index and the F-measure to assess the clustering results of a Gaussian mixture model based on expectation-maximization.

5. **Grading of prostate cancer: a work in progress**, Hystopathology, Oct. 2019.

   This paper describes the historical aspects of prostate cancer grading. Also summarise the current evidence for prognostic impact of each of Gleason grade subpatterns. They point out areas where such data are lacking and discuss potential implications of recent insights in prostate cancer grading for future improvement. Grading of prostate cancer has evolved substantially over time, not least because of major changes in diagnostic approach and concomitant shifts from late- to early-stage detection since the adoption of PSA testing from the late 1980s. After the conception of the architecture-based nine-tier Gleason grading system more than 50 years ago, several changes were made in order to increase its prognostic impact, to reduce interobserver variation and to improve concordance between prostate needle biopsy and radical prostatectomy grading. This eventually resulted in the current five-tier

grading system and they shows this with a much more detailed description of the individual architectural patterns constituting the remaining three Gleason patterns.

6. **Prostate Cancer Detection using Deep Convolutional Neural Networks** , Sci Rep 9, Dec. 2019.

   In this paper they introduced computer-aided detection (CAD) tools, diffusion-weighted magnetic resonance imaging (DWI) for accurate detection of prostate cancer. They found with deep convolutional neural networks (CNNs) significant success in computer vision tasks such as object detection and segmentation, different CNN architectures are increasingly investigated in medical imaging research community as promising solutions for designing more accurate CAD tools for cancer detection. In this work, their ultimate aim is to develope and implemente an automated CNN-based pipeline for detection of clinically significant prostate cancer (PCa) for a given axial DWI image and for each patient. DWI images of 427 patients were used as the dataset, which contained 175 patients with PCa and 252 patients without PCa. To measure the performance of the proposed pipeline they performed a test set of 108 (out of 427) patients were set aside and not used in the training phase.

7. **Prostate Cancer Detection Using Deep Learning and Traditional Techniques**, IEEE Access, Jan. 2020.

   This paper focus on deep learning methods to detect prostate cancer. This research employed deep learning long short-term memory (LSTM) and Residual Net (ResNet 101), independent of hand-crafted features, and is fine-tuned. The results were compared with hand-crafted features such as texture, morphology, and gray level co-occurrence matrix (GLCM) using non-deep learning classifiers such as support vector machine (SVM) Gaussian Kernel, k-nearest neighbor-Cosine (KNN Cosine), kernel naive Bayes, decision tree (DT) and RUS-Boost tree. This study reduces the features of carcinoma images, employed machine learning and deep learning approaches. For validation of training and testing data, a jack-knife ten-fold cross-validation method was used. The performance was measured using a confusion matrix such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy (AC), Mathews Correlation Coefficient (MCC), and area under the curve (AUC).

8. **Artificial Intelligence and Machine Learning in Prostate Cancer Patient Management—Current Trends and Future Perspectives** , Diagnostics, Feb. 2021.

   This paper focus on AI Based methods for prostate cancer detection. Artificial intelligence (AI) is the field of computer science that aims to build smart devices performing tasks that currently require human intelligence. Through machine learning (ML), the deep learning (DL) model is teaching computers to learn by example, something that human beings are doing naturally. They showed AI is revolutionizing healthcare. Digital pathology is becoming highly assisted by AI to help researchers in analyzing larger data sets and providing faster and more accurate diagnoses of prostate cancer lesions. They observed that, when applied to diagnostic imaging, AI has shown excellent accuracy in the detection of prostate lesions

as well as in the prediction of patient outcomes in terms of survival and treatment response. The enormous quantity of data coming from the prostate tumor genome requires fast, reliable and accurate computing power provided by machine learning algorithms. Radiotherapy is an essential part of the treatment of prostate cancer and it is often difficult to predict its toxicity for the patients. Artificial intelligence could have a future potential role in predicting how a patient will react to the therapy side effects. These technologies could provide doctors with better insights on how to plan radiotherapy treatment. The extension of the capabilities of surgical robots for more autonomous tasks will allow them to use information from the surgical field, recognize issues and implement the proper actions without the need for human intervention.

2.2 System Study Report

Prostate cancer is a most common type of cancer in men after the age of 50. If it is detected early it is curable. The variation of prostate cancer cells from normal cells is difficult to identify pathologist by looking into the images. It requires sufficient information about tissue patterns. Most of the studies takes only the gland structural and morphological information. It is important to consider nucleus and stroma composition of cells. If the gland part is high their is more chance to prostate cancer. There are several screening tests liks PSA, TRUS, DRE etc to diagnose prostate cancer but it has drawbacks. Recent advances in Computer-Aided Detection (CAD) using deep learning have brought the immense scope of automatic detection and recognition at very high accuracy in prostate cancer. The proposed work proposes a segmentation method to segment histopathology image of prostate cancer and a deep learning approach for classification of Gleason patterns.

Chapter 3

Project Design

This work proposes image processing techniques and deep learning techniques to segment and classify prostate cancer by analysing histopathology images. Histopathology images are created by obtaining a tissue sample from a prostate biopsy and treats it by a staining protocol to highlight the histological structures. This work aims to grade prostate cancer from prostate histopathology images using a CNN based network. Information about glandular structures is considered for grading prostate cancer. In this work morphological features are well analyzed for determining how advanced your prostate cancer is. The project work consists of three steps they are (1) Pre-processing (2)Gland segmentation (3)Classification.

3.1 Resource Requirements

3.1.1 Hardware & Software Requirements

Operating System	: Any Operating System
Supporting Software and Libraries	: Python, Tensorflow, Keras, OpenCv
Graphics Card	: 6GB NVIDIA GeForce GTX 1060 GPU
Processor	: Intel Core i5 10th Gen
RAM	: 8GB
Supporting Environment	: Google Colab

3.1.2 Dataset

All data and informations regarding prostate cancer is obtained from Regional Cancer Centre, Thiruvananthapuram. The dataset used in this work consists of three classes of histology images. They are grade 3, grade 4 and grade 5. The dataset is divided into training set, testing set and validation set. Since CNN network needs large number of dataset for training, images are augmented to generate new images of same quality and size without lossing any informations.

3.2 Methodology

The proposed work is developed with three steps which include pre-processing, gland segmentation and classification(grading). Prostate cancer of grades 3, 4 and 5 are collected and prepared

dataset. In the pre-processing step the image is resized and enhanced using median filter. Principal component analysis and k-means clustering method is applied to segment gland parts from histology images. Grading is performed on segmented images using a CNN based network called MobileNet.

3.2.1 Pre-processing

Pre-processing an image is a must so that programs work properly to give the expected output. The aim of pre-processing is to improve the quality of the image so that we can analyse it in a better way. By pre-processing we can suppress undesired distortions and enhance some features which are necessary for the particular application we are working for. Those features might vary for different applications. To train a network and make predictions on new data, the images must match the input size of the network. In order to adjust the size of images to match the network, images must be resized or scaled.

3.2.2 Segmentation

Image segmentation is the process of partitioning a digital image into multiple segments. Medical image segmentation refers to the process of extracting the desired object from a medical image. The goal of segmenting data is to identify areas of anatomy required for particular study. Segmentation allows more precise analysis of anatomical data by isolating only necessary areas.Segmenting in histology images aims to emphasize tissue structures for detection and diagnosis of cancer. Histology segmentation possess high worth in disease diagnosis and grading.

The segmentation method utilized in this work is a combination of two methods. They are Principal component analysis (PCA) followed by K-means clustering method. These methods segment tissue components such as stroma, lumen and cytoplasm.

Principal Component Analysis (PCA)

Principal component analysis is a statistical tool applied on digital images for image analysis, pattern recognition and image compression. It is an image segmentation task in the compressed domain or low dimension domain, which is performed to overcome some limitations such as inefficient pre-processing methods and limited storage for high resolution images. The working of PCA is based on properties of eigen values and eigen vectors of matrices.

K-means Clustering

K-means clustering algorithm is an unsupervised algorithm and it is used to segment area of interest from background. It clusters or partitions the given data into K-clusters or parts based on K-centroids. This method clusters the components of image until specified criteria is reached.

3.2.3 Classification(Grading)

The classification algorithm is a supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from

the given dataset or observations and then classifies new observation into a number of classes or groups.

In this work, MobileNetV2 is used to perform classification or grading of prostate cancer. Classification is performed in two phases. They are training and testing. Classification can be performed with deep learning or non deep learning techniques. The process of staging of cancer is called grading. Gleason grading is well known method pathologist used to grade prostate cancer. The dataset contains Gleason grade patterns of grade 3, 4 and 5. In deep learning techniques the features of training data is automatically learned by the model and predict the category of new data. But in non deep learning technique the features need to be extracted first and inputted to the model along with dataset for learning. Computer assisted systems using Deep learning techniques gives better result than non deep learning techniques in grading prostate cancer.

MobileNet are efficient models for classifying medical images. They are a class of convolutional neural networks that can be used for image recognition and classification. They are small, low latency models. Because of their small size, these are considered great deep learning models. MobileNet has a streamlined architecture that uses depthwise separable convolutions to construct lightweight deep convolutional neural networks. It significantly reduces the number of parameters when compared to the network with regular convolutions with the same depth in the nets.

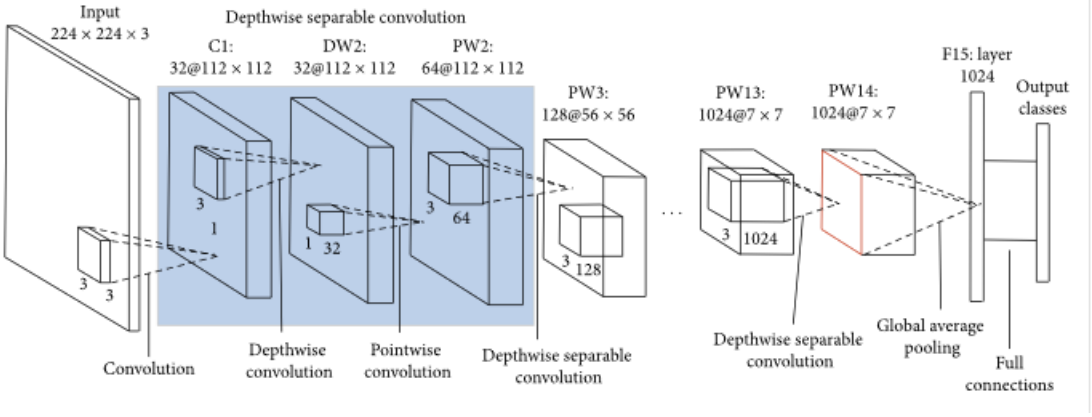

Figure 3.1: Architecture of MobileNet

3.2.4 Block diagram

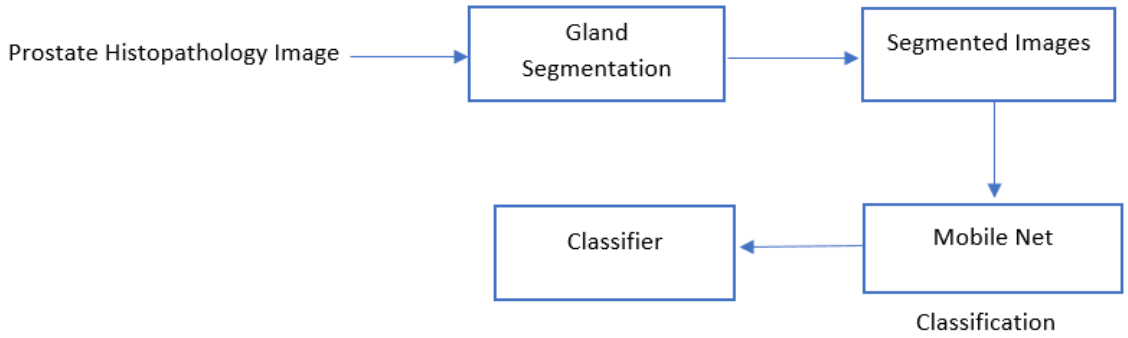

Figure 3.2: Block Diagram of Training Phase

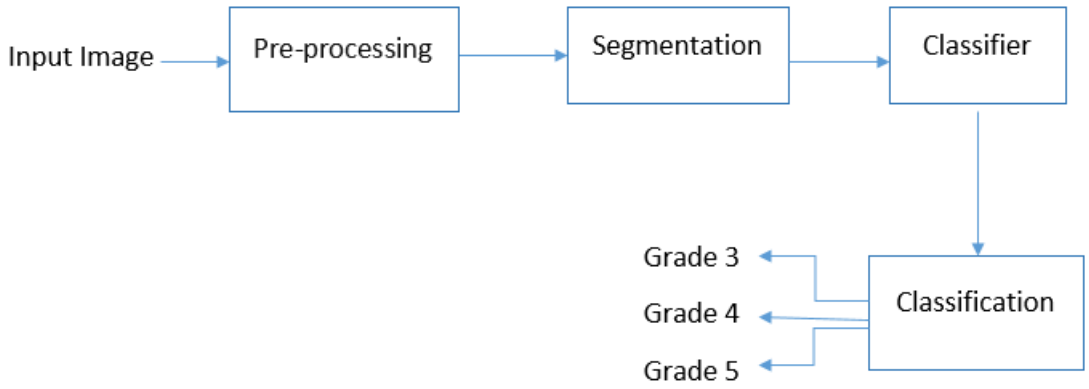

Figure 3.3: Block Diagram of Testing Phase

Chapter 4

Implementation

The project work is initialized by preparing the dataset for three Gleason patterns of prostate cancer: grade 3, 4 and 5. Since the images on each class is limited, they are augmented to about 250 new images for each class. Then divided the new dataset into training set, validation set and testing set.

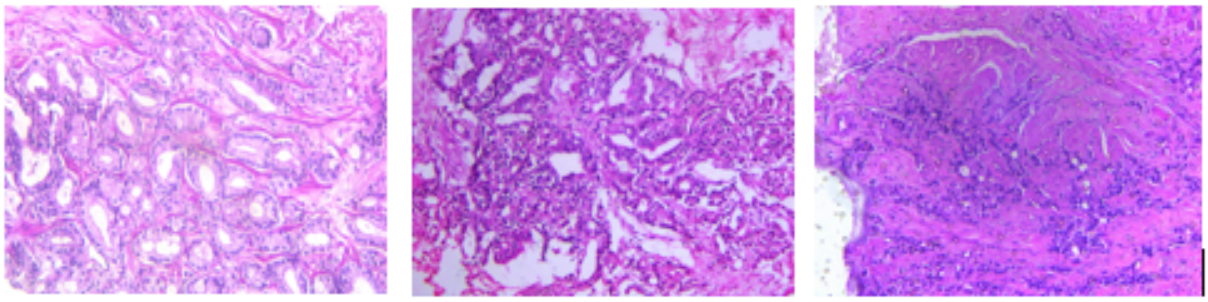

Figure 4.1: Sample Images of Grade 3, 4 and 5 From Dataset

4.1 Pre-processing

Pre-processing is performed in testing phase of the model. The images in the dataset are already pre-processed. In testing, the new image data is first applied with median filter for enhancement and noise removal. Also images are resized to 224×224, since the specified image size of MobileNet is 224×224.

4.2 Gland Segmentation

PCA and K-means clustering is performed to segment gland components such as stroma, lumen and nuclei. First images are applied with PCA to emphasize the componets of glands without lossing the informations and quality of images. Then K-means clustering is performed by giving K value = 4, that clusters the image into 4 components. Each category of images are separately

segmented and saved in a new folder. These segmented images are given to MobileNet for training.

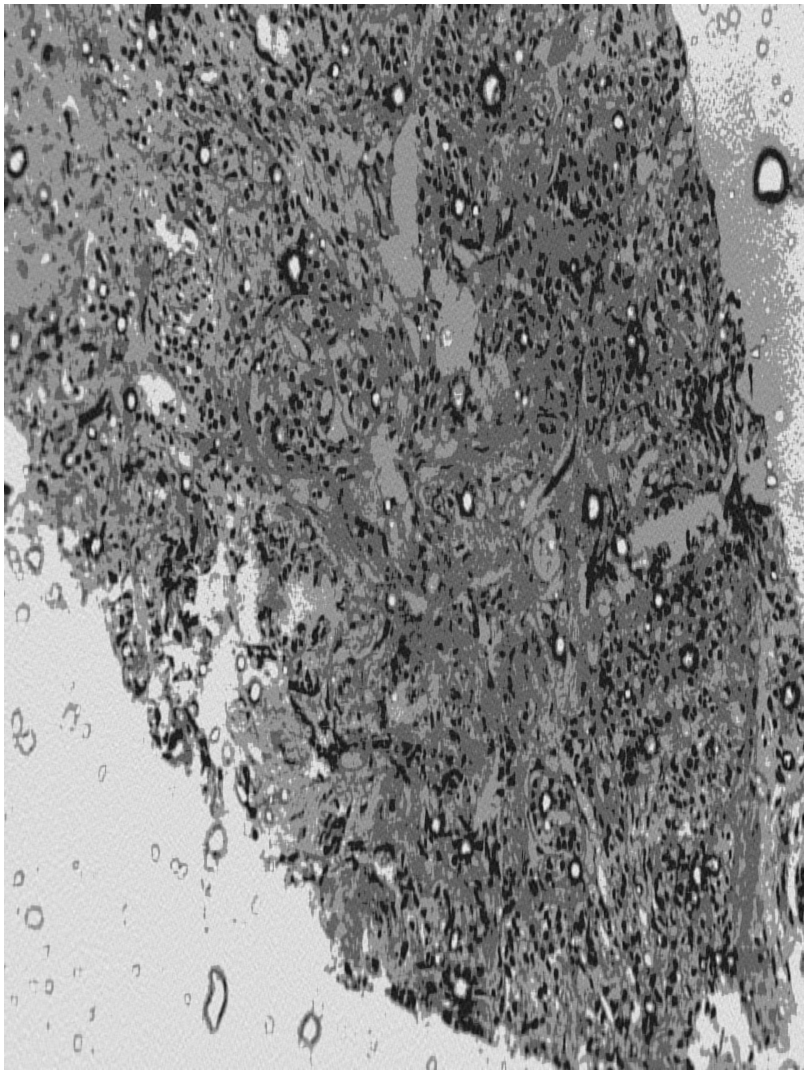

Figure 4.2: Segmentation Using Principal Component Analysis(PCA)

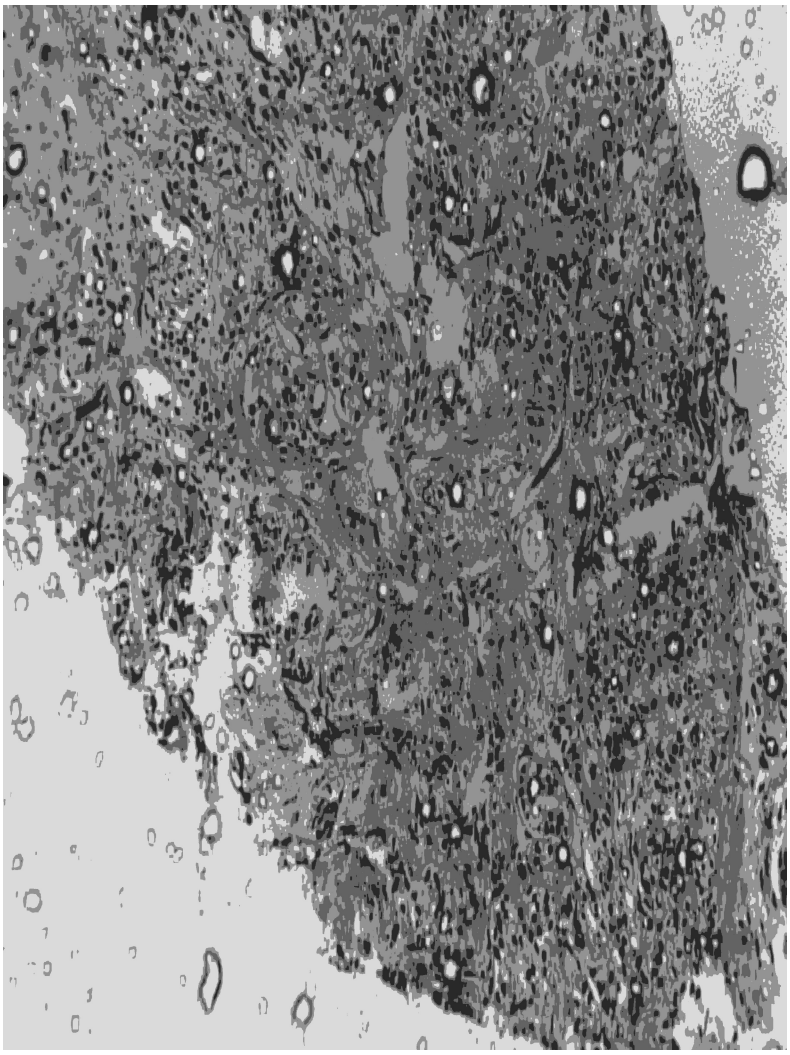

Figure 4.3: Segmentation Using Principal Component Analysis(PCA) Followed by K-means Clustering

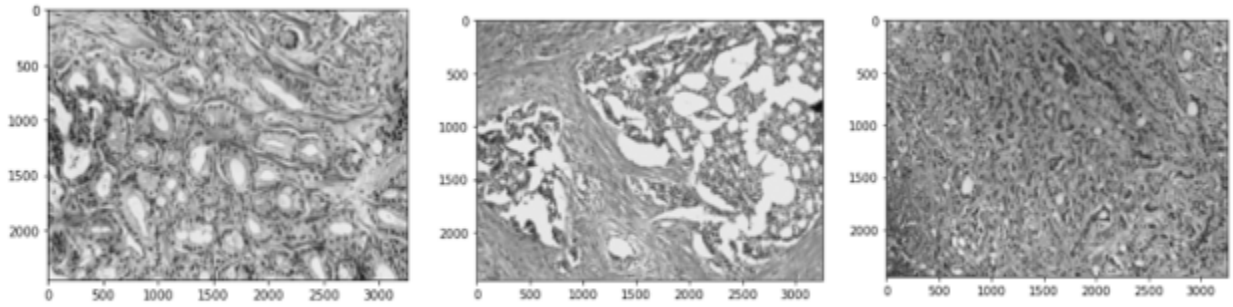

Figure 4.4: Segmented Images of Grade 3, 4 and 5

4.3 Classification

Classification is performed in two phases. They are training and testing. In training phase, the pre-processed images from training set is first segmented and applied the segmented images to MobileNetV2. The input to MobileNet is training and validation set. The network is trained in 100 epochs.

In testing phase, MobileNet is tested with a test set of segmented images and obtained testing accuracy of 93%.

4.4 GUI

This work is implemented as a desktop application. The user interface for the particular system is figure out below. The button, upload image in the window is to select the image from the computer system for testing. Pre-process button will perform pre-processing of the image that is uploaded. Segmentation button is to perform segmentation operation in the pre-processed image. After segmentation grade is predicted using the classifier by clicking predict grade button. Figure shows the window opened for testing the model. An input image from test set is uploaded, it is pre-processed then segmented and predict the class label as grade 3.

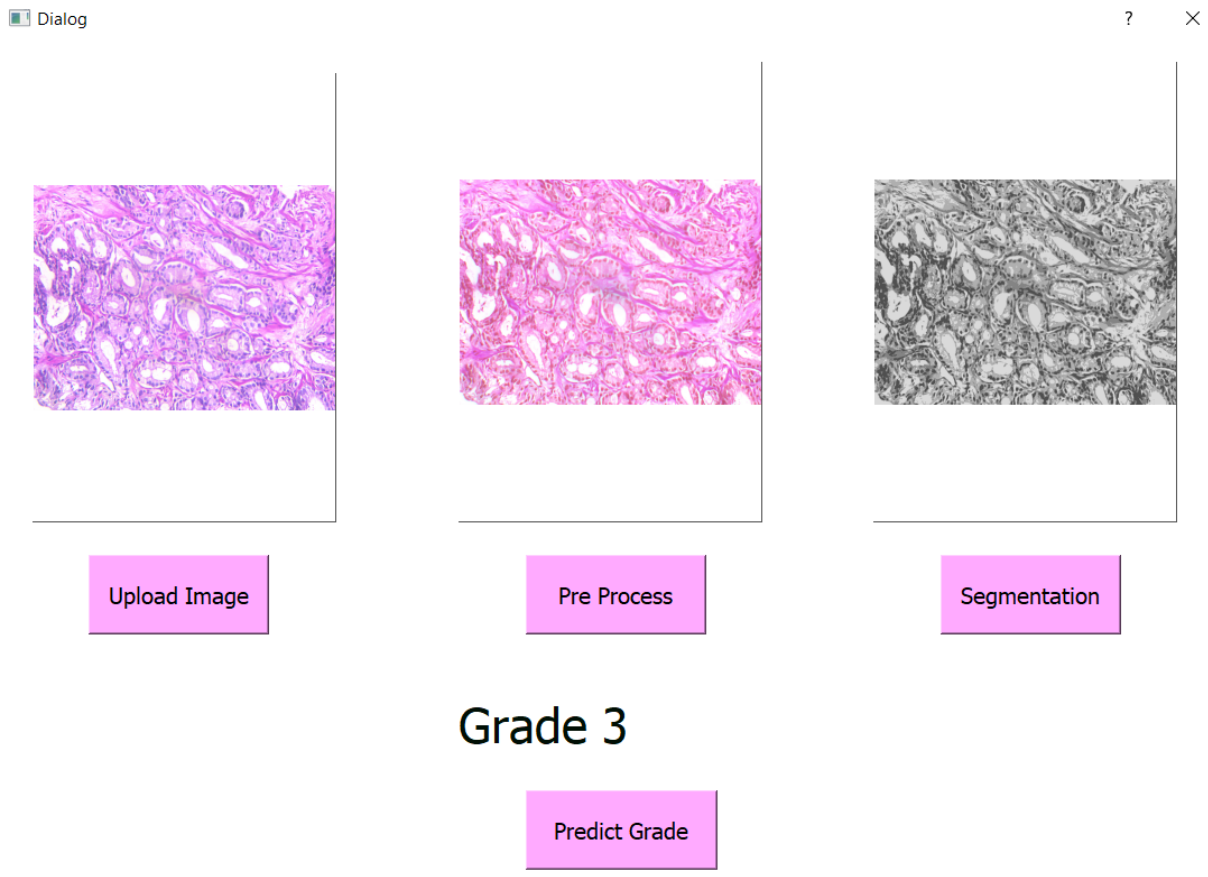

Figure 4.5: User Interface

Chapter 5

Results & Conclusions

5.1 Performance Analysis

The performance of the model is analyzed in both training and testing phase. In training phase accuracy and loss of the model is plotted. It is found that the accuracy of the model is above 0.95 and the loss of the model is very small in each iteration. Once the model is trained it is tested with a test set of samples. To evaluate the model in testing phase, a confusion matrix is a best tool. A confusion matrix is a summary of prediction results on a classification problem. Calculating a confusion matrix can give you better idea of what your classification model is getting right and what types of errors it is making. In this work, there are three classes or grades. Various parameters associated with confusion matrix are true positive(TP), true negative(TN), False positive(FP) and false negative(FN). Based on these parameters, precision, recall, specificity and accuracy of the model is calculated. The accuracy obtained for proposed method is 93%

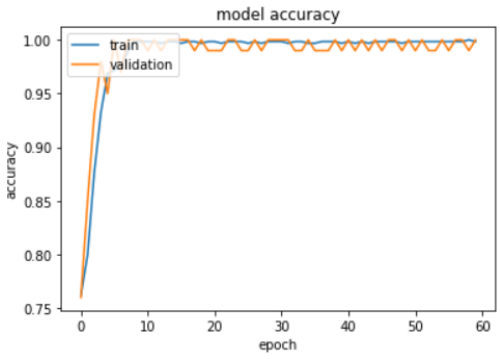

Figure 5.1: Model Accuracy

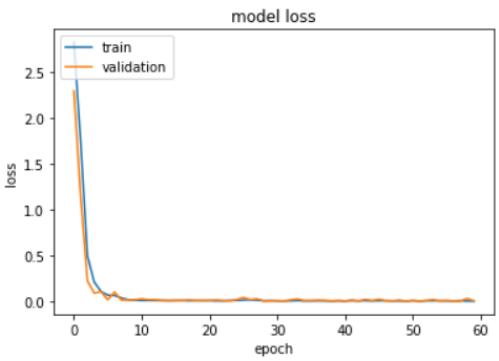

Figure 5.2: Model Loss

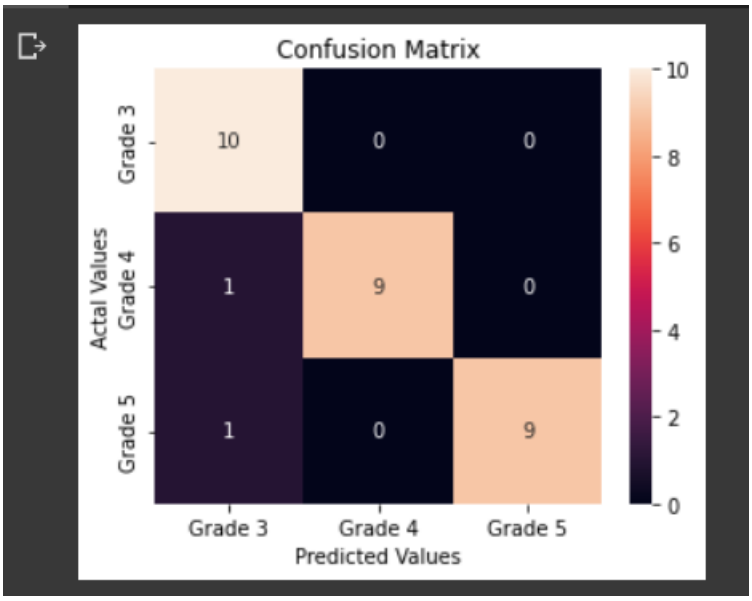

Figure 5.3: Performance Analysis

	precision	recall	f1-score	support
Grade 3	0.83	1.00	0.91	10
Grade 4	1.00	0.90	0.95	10
Grade 5	1.00	0.90	0.95	10
accuracy			0.93	30
macro avg	0.94	0.93	0.93	30
weighted avg	0.94	0.93	0.93	30

Figure 5.4: Classification Report

5.2 Conclusion

The project aims to propose and develop a model that grades prostate cancer into three main grades. From the study, recently many works were conducted on prostate cancer detection and grading using different imaging modalities like MRI, CT images but analyzing histopathology images is found difficult. Also found that gleason grading system is a well known system to evaluate the prognosis of men with prostate cancer. So this work focus on grading of prostate cancer from histopathology images using gleason grading system. The objective of this project is to implement a deep learning model that is able to grade prostate cancer accurately, thereby analyse the severity of the cancer. The project starts with collecting histopathology image dataset of prostate cancer, then image pre-processing is performed to enhance the images. These images are then segmented using Principal component analysis and K-means clustering. Grading of prostate cancer is done with a CNN network, MobileNet. This model achieved 93% accuracy on test dataset.

5.3 Future Scope

In future, the proposed method can be extended for predicting Gleason scores from 6 to 10 by using patches of full slide images of prostate histopathology.

References

[1] Li Zhang, Longchao Li, Min Tang, Yi Huan, Xiaoling Zhang and Xia Zhe, " A New Approach to Diagnosing Prostate Cancer Through Magnetic Resonance Imaging", Alexandria Engineering Journal, Elsevier, vol 60, Issue 1, February 2020.

[2] Saqib Iqbal, Ghazanfar Farooq Siddique, Amjad Rehman, Lal Hussain, Tanzila Saba, Usman Tariq and Adeel Ahamed Abbasi, "Prostate Cancer Detection Using Deep Learning and Traditional Techniques", IEEE Access, vol 9, Feb 2021.

[3] Jimmy C Azar, Christer Busch, Ingrid B Carlbom, "Microarray Core Detection by Geometric Restoration", Analytical Cellular Pathology, 2012.

[4] Yuchun Li, Mengxing Huang, Yu Zhang, Jing chen, Haixia xu, Gang Wang and Wenlong Feng, "Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer", IEEE Access, July 2020.

[5] Jimmy C Azar, Christer Busch, Ingrid B Carlbom, "Histological stain evaluation for machine learning applications", International conference on Medical Image Computing and Computer Assisted Intervention, oct 2012.

[6] C F Kweldam, G J van Leenders and T van der Kwast, "Grading of Prostate Cancer: A Work in Progress", Histopathology, Oct 2019.

[7] Sunghwan Yoo, Isha Gujrathi, Masoom A. Haider and Farzad Khalvati, "Prostate Cancer Detection Using Deep Convolutional Neural Networks", Sci Rep 9, Dec 2019.

[8] Saqib Iqbal, Ghazanfar Farooq Siddique, Amjad Rehman, Lal Hussain, Tanzila Saba, Usman Tariq and Adeel Ahamed Abbasi, "Prostate Cancer Detection Using Deep Learning and Traditional Techniques", IEEE Access, vol 9, Feb 2021.

[9] Octavian Sabin Tătaru, Mihai Dorin Vartolomei, Jens J. Rassweiler, Oșan Virgil, Giuseppe Lucarelli, Francesco Porpiglia, Daniele Amparore and Matteo Manfredi, "Artificial Intelligence and Machine Learning in Prostate Cancer Patient Management—Current Trends and Future Perspectives", Diagnostics, Feb 2021.

[10] Singhal, N., Soni, S., Bonthu, "A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies", Sci Rep 12, 2022. https://doi.org/10.1038/s41598-022-07217-0

[11] Guy Nir, Ladan Fazli, Peter C Black, Larry Goldenberg, Septimiu E Salcudean, "Deep Learning-Based Gleason Grading of Prostate Cancer From Histopathology Images-Role of Multiscale Decision Aggregation and Data Augmentation", IEEE J Biomed Health Inform. 2020 May;24(5):1413-1426. doi: 10.1109/JBHI.2019.2944643.

[12] M., Saha, A., Brand, "Deep learning–assisted prostate cancer detection on bi-parametric MRI: minimum training data size requirements and effect of prior knowledge", Imaging Informatics and Artificial Intelligence – Eur Radiol 32, 2224–2234 (2022). https://doi.org/10.1007/s00330-021-08320-y.

[13] Armato, S. G., Petrick, N. A. Drukker, K.Prostatex, "Prostate MRi classification challenge", (conference presentation). Proceedings of the SPIE, Volume 10134, id. 101344G 1 pp.(2017).134 (2017).

[14] Liu, S., Zheng, H., Feng, Y. Li, "Prostate cancer diagnosis using deep learning with 3d multiparametric MRI", In Medical Imaging 2017: Computer-Aided Diagnosis, vol. 10134, 1013428 (International Society for Optics and Photonics, 2017).

[15] Fehr, "Automatic classification of prostate cancer gleason scores from multiparametric magnetic resonance images", Proc. of the Natl. Acad. of Sci. 112, E6265–E6273 (2015).