Automatic Prostate Cancer Grading Using Deep Architectures

Author
Muhammad Mohsin
NUST CEME 00000206739

Supervisor
Dr. ARSLAN SHAUKAT

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

July, 2021

Automatic Prostate Cancer Grading Using Deep Architectures

Author

Muhammad Mohsin
NUST CEME 00000206739

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Computer Engineering

Thesis Supervisor

Dr. ARSLAN SHAUKAT

Thesis Supervisor‘s Signature: _____

DEPARTMENT OF COMPUTER ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

July, 2021

Declaration

I, Muhammad Mohsin, the author of this thesis, hereby declare that this thesis entitled "Automatic Prostate Cancer Grading using Deep Architectures", is solely my own work which has been submitted to fulfil a partial requirement for the award of degree of Masters in Computer Engineering at NUST College of EME Rawalpindi. No such piece of writing has been submitted previously for any other award or at any other university. Wherever others' contribution is involved, whether it be a piece of information derived from others' work or a collaborative research work or discussion, it is acknowledged explicitly with references in the text. The work is complied with the University's requirements and rules of practice for the research degree programs. I understand that, once accepted, this thesis will belong to the university's ownership and I will honor university's norms in this regard.

Signature of Student

Muhammad Mohsin

NUST CEME 00000206739

Language Correctness Certificate

The thesis has been read thoroughly by an English language expert and is free of any language error, like typing, syntax, semantic, grammatical, and spelling mistakes. Moreover, the format given by the University for the Thesis of MPhil degree program, is followed in the document.

Signature of Student

Muhammad Mohsin

NUST CEME 00000206739

Signature of Supervisor

Dr. ARSLAN SHAUKAT

Copyright Statement

- Copyright in text of this thesis rests with the student author. Copies (by any process) either in full, or of extracts, may be made only in accordance with the instructions given by the author and lodged in the Library of NUST College of E&ME. Details may be obtained fromthe Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.

- The ownership of any intellectual property rights, which may be described in this thesis, is vested in NUST College of E&ME, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the College of E&ME, which will prescribe the terms and conditions of any such agreement.

- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST College of E&ME, Rawalpindi.

Acknowledgements

First and foremost, I praise and thank Allah (S.W.T), the most Gracious and the entirely merciful, for giving me the strength, guidance, and opportunity to undertake this research project and completing it successfully. Indeed, we all ultimately depend on Him for guidance, support, and subsistence.

A very special and sincere gratitude goes out to my research supervisor, Dr. Arslan Shaukat, Tenured Associate Professor and Head of PG Program, at NUST College of EME Rawalpindi, whose invaluable guidance and encouragement enabled me to complete this thesis. I am deeply inspired by his wisdom, diligence, integrity, and motivation. He consistently steered me in the right direction by giving constructive feedback and useful comments on my work. Working under his supervision was truly a great privilege for me. I am indebted to him for sharing his expertise, showing empathy in difficult times and offering a strong affectionate relationship. I always felt appreciated during our interactions which helped improve my performance and confidence. I acknowledge that any error in this thesis would be my own and should not stain the good repute of my respected supervisor.

I will forever be grateful to rest of GEC team members, Dr. Muhammad Usman Akram and Dr. Farhan Hussain, for their endless and genuine support during the whole time. I truly appreciate their patience while helping me overcome numerous barriers. Their insightful comments during our discussions provided me many learning opportunities and helped keep my thesis harmonious. I hope I have been able to address most of the points raised by them on my work. I also want to thank them for making my defense an enjoyable moment.

I am profoundly grateful to my parents and siblings for their prayers, unconditional love, protection and amazing support for developing my career. They really understood challenges of my study and provided spiritual and emotional support along the way. I could never have accomplished this success without their continuous support. Finally, a very humble thank to all those individuals who have supported me by any means to complete my thesis.

Muhammad Mohsin Khan

In dedication

To my beloved father, ***Noor Samad ,*** *for encouraging and supporting*

To my beloved mother, ***Mrs. Noor,*** *for never giving up on me and pushing me to be my very best*

Abstract

Prostate cancer is the second most aggressive type of cancer diagnosed in men, seriously affecting people's life and health. Prostate cancer detection and grading in advance is a very critical step for pathologists. Large scale inter observer reproducibility exists in staging the prostate biopsies which leads us to move towards a Computer based model, which could accurately detect and grade the cancerous prostate. Due to recent development in the field of digital pathology, tissue microarrays (TMA) images are generated from whole slide images resulting in less computational procedures and achieve good performance. This thesis is focused on deep learning model to automatically stage the cancer instead of feature engineering based models Deep learning models directly learn the features via convolutional layers and achieve good accuracy as compared to feature engineering based models. We have used two datasets, Harvard dataset and Gleason Challenge 2019, for implementation of our proposed model. Our proposed UNET based architecture is used for training as well as testing and evaluation. We have used different deep learning models for our UNET based encoder and achieved 0.728 and 0.732 average Cohen's kappa with F1 on both datasets respectively. The results show that our proposed UNET based deep learning model performs better as compared to other state of art models. Hence, it would help pathologists to automatically grade the prostate biopsies with high accuracy.

Table of Contents

List of Figures

List of Tables

1 Chapter 1: Introduction

1.1 Brief Description

Prostate cancer is the second most deadly form of cancer commonly found in men in the United States [1]. In developed countries, prostate cancer is increasing exponentially due to high living standards and population explosion. A very large number of prostate cancerous patients die every year due to insufficient diagnosis environment and large scale inter observer variation between pathologists. This all led to design a model of early detection and correctly classify the cancerous grade. Since 1960, the Gleason grading algorithm has been the most reliable and effective prognostic predictor for prostate cancerous cells [2]. The Gleason scoring algorithm is highly recognized by the World Health (WHO), and it was updated and revised by the International Society of Urological Pathology in 2005 and 2014 (ISUP) [3]. Although there have been several advances in the clinical diagnosis of prostate cancerous cells, Gleason scoring dependent on histology remained the most effective prognostic indicator of prostate carcinoma's early diagnosis and staging [4].

The architectural pattern of cancerous cells is used to render the histological evaluation. The architecture pattern consists of well differentiated cells to the poor ones. The Gleason score is totally based on this architectural pattern. **Figure 1.1** shows micro-array images which have been classified into benign and Gleason score 6 to 8, depending on their architectural patterns.

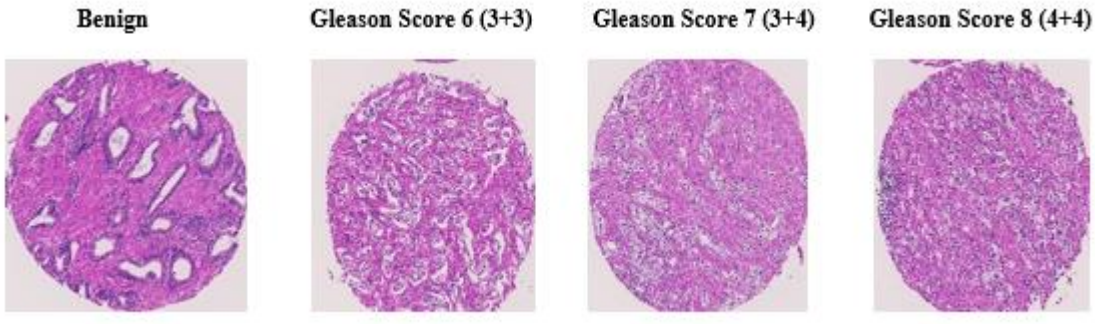

Figure 1.1: Example of tissue microarrays images (TMA), benign to Gleason Score 8 [5]

The Gleason score assigned is totally based on pathologist's review and it's a time-consuming but essential phase with a lot of inter- and intra-observer variation. This problem specially occurs when differentiating Gleason grade 3 from Gleason grade 4 having Gleason score equal to 7 (4+3 or 3+4), which has a very complex impact on further treatment. Since major medical decisions are based on rating of biopsy specimens, dire need of automatic prostate cancer grading model is obvious [6-8]. There are many feature engineering based techniques used for automatic Gleason grading [9]. The success of feature engineering based techniques is totally dependent on how accurately features are extracted and their compatibility with model. In prostate cancer grading, due to high correlation between cancerous cells to the healthy cells, there are high chances of misclassification which leads to the failure of automatic computer-based models.The computer-aided method of identification focused on convolutional neural network has always been considered to play a significant role in medical image analysis [10, 11]. This method has replaced the traditional way of drawing out features for image classification with a totally different approach of allowing the computer to finalize which features have to be considered. The outstanding results on standard datasets have made CNN a widely used technique for pattern identification. Therefore, we worked on CNN system to analyze different tissue microarray images and assign them Gleason score.

1.2 Aim of the Thesis

This thesis is aimed on executing a classifier on a limited dataset of (H and E) Tissue Micro Array (TMA) images and finding out how Convolutional Neural Network (CNN) technique can be employed on it to get positive outcomes. Besides, the limited dataset convinced us to use various augmentation practices as well. As a whole, our research shows that the systems based on deep learning approach have great potential of producing accurate and reproducible results, particularly in case of heterogeneous structures.

1.3 Scope of the Thesis

We intend to develop a system that could easily be used by physicians without getting involved into the complex algorithms, working in the background. The development of such a system, offering masterly performance, for evaluating prostate biopsies would encourage Pakistani

specialists to use this automated system for cancer diagnosis. This would save pathologists' valuable time since they have to go through significant number of slides, often involving further immune staining, to reach a decision. Besides, pathologists have to measure various parameters (such as length, mitotic rate, surface areas etc.) for ordinary evaluating systems [12]. In addition, there is a great deficiency of oncologists in Pakistan, with the ratio of medical oncologists to population being very low. Hence, such software would decrease the burden on expert specialist by offering a balance between prognostication and practicability in everyday clinical practices.

1.4 Thesis Problem Statement

In this project, the assessment of tissue microarrays and metadata triggered the verbalization of problem statement. The intention was to train a machine learning program by using tissue microarray images which could be utilized afterwards to find out seriousness of disease on the basis of Gleason score assigned to unviewed biopsies. This is a multiclass classification task since five distinct grade groups are involved. Multiclass classification is complicated, especially due to class variance and small dataset, as compared to binary classification of clinical importance. Besides, a small set of data was available, which was an additional problem, for testing and training which was addressed by applying suitable augmentation techniques. The success of the software depends on how effectively features are extracted from biopsy images and how accurately the Gleason scores are assigned to specimens. Hence, the key challenge in classifying prostate cancer specimen is to develop a good feature extraction technique and execute an effective mask predicting classifier.

1.5 Method

In this dissertation, we have discussed in detail the latest Prostate Cancer Grading techniques which are being used in clinical process nowadays. We are working on MICCAI Gleason Grade Challenge 2019 Dataset and publically available Harvard dataset which have both training as well testing data. The Dataset consists of Tissue Micro Arrays (TMA) of prostate biopsies. TMA images are more useful as compared to whole slide images because they contain only the cancerous region of image which directly reduces the computation part. As deep learning models need larger number of data for their training, data augmentation techniques are applied for

better training. UNET Model has used pixel wise classification of images and predicts the images on test data. On both datasets, our CNN-based UNET model achieved state-of-the-art performance with scores of 0.728 and 0.732, respectively. The system's performance was evaluated on (245) biopsy specimens, which were graded by (6) expert urologists independently from ISUP. By evaluating the extent of cancer as calculated by the method, Cohen's kappa combined with F1 score was used to calculate the concordance between the grades reported by expert urologists and those assigned by the automated computer system.

1.6 Circumscription

Tissue images obtained from different sources may differ in appearance, something that will not be discussed in this study. In order to study dimensions of small histological images, magnification of the order of (40X) will be done and the size of the related image crops will be delimited. No research will be performed on the impact of contrast or color exploitation.

1.7 Structure of the Thesis

The thesis is divided into the following six chapters.

Chapter 1 describes the introduction, aim and scope of the thesis with problem statement, Chapter 2 describes the background for an appropriate understanding of anatomy of prostate, prostate cancer and its types, staging and grading of prostate cancer and screening techniques.

Chapter 3 discusses the previous techniques utilized for image pixel level classification in order to explore the most effective method for image analysis. The discussed techniques include best deep learning model architectures.

Chapter 4 is based on the proposed methodology and algorithm for segmentation and extraction of cancerous regions. Then, Chapter 5 contains the results. The proposed algorithm has been statistically shown to be a reliable method for prostate cancer diagnosis. The conclusions are also compared with manual grading of expert pathologists. Pixel-based approach has been used to evaluate the suggested algorithm.

In chapter 6, conclusion, contributions, and recommendations for further work have been presented.

2 Chapter 2: Related Background

2.1 Tumor

Normally, the cells in a human body grow and divide in an order. Every cell in a body performs a certain job and when old cells are destroyed, new cells take their place and the cycle goes on. But when cells divide in an uncontrolled way and do not die, they form an unnecessary mass of tissue called tumor. The tumor keeps on growing if more and more cells continue to accumulate in the mass. While some tumors are benign, others are malignant.

2.2 Benign Tumors

Benign tumors are noncancerous and their spreading to other areas of the body is not possible. They usually do not grow back once removed. These tumors mostly do not cause any harm but in some cases, if they press against vital structures like blood vessels or nerves, they may result in pain or some other problems.

2.3 Malignant Tumors

These tumors are cancerous, and they can spread to other areas of the body easily. The cancerous cells contain abnormal chromosomes and DNA, having large and dark nuclei. These cells may grow back after removal and require aggressive treatment.

There are different types of malignant tumors which arise in different types of cells. Major types include carcinomas, sarcoma, melanoma, lymphoma and leukemia. Sarcoma arises in multiple locations in the body like blood vessels, bones, muscles and in soft and connective tissues. Melanoma develops in cells that produce pigment in skin while lymphoma arises in the immune system and leukemia develops in blood tissues.

Among all these types, carcinomas are the most common types which start in epithelial tissue. Epithelial tissues are the major tissues found in glands and form the covering of internal organs, such as the liver or kidneys. There are different subtypes of carcinoma cancer depending upon the type of epithelial cells. Some common types include,

1. Tumors of the squamous cells
2. Adenocarcinoma
3. Carcinomas of the transitional cell
4. basal cell carcinoma

Our study is concerned with adenocarcinoma of the prostate. Adenocarcinoma basically develops in glandular cells called adenomatous. These cells are responsible for releasing fluids into the body or excreting some substance from the body. Most cancers in prostate are adenocarcinomas. About 95% of prostate cancers are adenocarcinomas [13].

2.3 Anatomy of Prostate Cancer

2.3.1 Prostate Gland

Prostate is a tiny gland in men's reproductive system that is around the size of a walnut. **Figure 2.1** shows its location and structure which is present above the males' reproductive organ (penis) and below the bladder. The urethra, which takes urine from the bladder out of the body, flows along the middle of the prostate. In the urethra, the prostate secretes a seminal milky fluid. This fluid combines with sperm and forms a part of semen to assist in its nourishment and transportation out of the body.

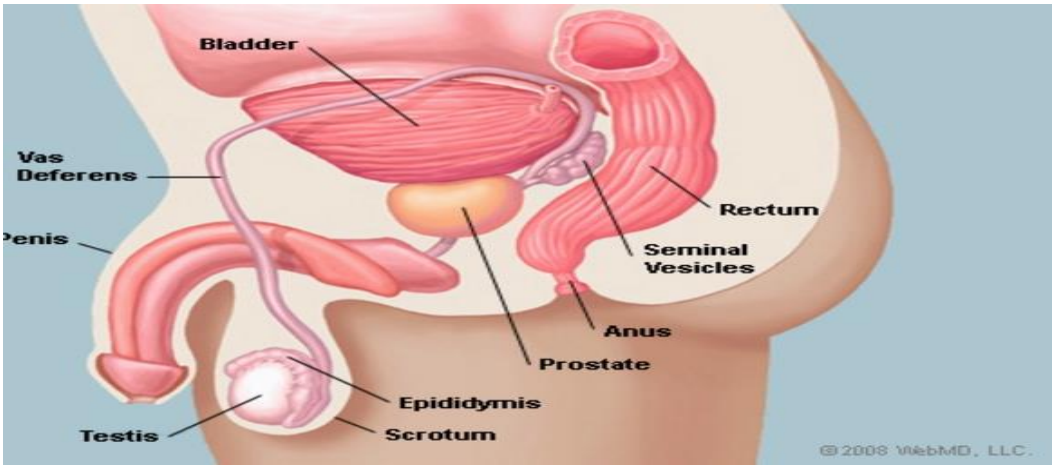

Figure 2.1: Picture of the Prostate[14]

2.3.2 Prostate Cancer

Prostate cancer is said to arise as prostate cells multiply out of control. Prostate cancer is one of the most common malignant tumors, and it is the sixth leading cause of death in men around the world [15].

Prostate cancerous cells can spread to other areas of the body, where they can develop a new tumor. Prostate cancer metastasis occurs when cells from the prostate tumors break apart and migrate to other areas, most commonly the bones and lymph nodes, causing damage to healthy tissues. Since the new affected region includes the same type of irregular cells as the initial cancer, the new tumor is assigned the same name. For instance, when prostate cancer cells migrate to the bones, the condition is considered metastatic prostate cancer rather than bone cancer.

2.3.3 Symptoms

Prostate cancer doesn’t usually show any symptoms until the prostate gets large enough to press against the urine carrying tube (urethra). When urethra is pressed, following symptoms can be noticed.

- Change of urination pattern such as pain while urinating, more frequent desire to urinate and having a feeling that your bladder is not fully emptied.
- blood in the semen

2.4 History of Prostate Cancer

Prostate cancer's structure and experience have resulted in highly critical factors in treatment, detection, and diagnosis. Changes in the DNA, which provides orders to the cells to replicate, are believed to be the origin of prostate cancer, according to scientists. One of the major causes of prostate cancer, according to studies, is an individual's diet and age, but it also has a lot to do with tumor growth in the prostate gland. Furthermore, genetics play a major role in the growth of prostate cancer and must be taken into account. Since prostate cancer is the most common cause of cancer death in men above a certain age, all men around the world should be forced to take preventative medicines. Prostate cancer is primarily characterized by changes in body cells and the physiology making its diagnostic approaches achievable through cell sorting

and grading. Cell grading approach is helpful in identifying affected cells through the employment of various approaches, in our case the use of deep architectures, a powerful tool that when well harnessed can be of great help in ensuring that prostate cancer management is easy and efficient. Prostate cancer management and care is largely focused on detecting infected cells, assessing cancer stage, and eventually determining the extent of infection. With this knowledge, a proper treatment plan may be taken to ensure that the patient gets the required drug to ensure success [16].

Over the years, scientists have produced a number of prostate cancer therapies that have proved to be reliable. Prostate cancer is more prevalent in men of 50 years and older, according to studies. While the cause of prostate cancer is unclear, a variety of risk factors have been attributed to the condition, including age, African-American ethnicity, a high-fat diet, family background and, perhaps most interestingly, exposure to rubber or cadmium. Men of all ages must know the prostate cancer prevention, treatment choices, screenings and, most notably, diagnosis. According to the American Cancer Society, "about 317,000 new cases of prostate cancer are diagnosed each year in the United States, with approximately 41,000 men dying each year." Prostate cancer can affect men all over the world at any stage in their life, whether it is them or a close friend. As a result, having the necessary medications and vaccines would be sufficient in treating prostate cancer.

While there is no one-size-fits-all solution for reducing prostate cancer, there are a variety of steps that can be done to reduce the possibility of developing it. The most important way to avoid prostate cancer is to eat well and follow a safe lifestyle. In their daily lives, most men eat so much meat and not enough healthy foods like vegetables and fruits. Consumption of red meat and dairy products is favorably correlated with P.C.A., while consumption of olive oil and green tea is negatively associated. This may result in speeding the progression of prostate cancer in future. Furthermore, it has been discovered that eating fish rich in fatty acids, such as tuna, herring, and mackerel, is the safest way to avoid prostate cancer. Although scientists are unsure about the reason, eating fish oils tends to inhibit tumor development in the prostate gland. "Men who consume these types of fish three days a week and complement their diet with some form of marine fatty acid supplement would have a much lower percentage risk of developing prostate cancer," according to a study ("Patient's Guide"). It has also been established in other nations, such as Asia,

that drinking green tea and soy decreases the risk of contracting prostate cancer. Vitamins D and E are both taken into account to ensure the body's health. Men should take the necessary vitamins and nutrients for their bodies so that they do not grow or to risk prostate cancer when they get older.

The signs of prostate cancer must be recognized, as well as the stage at which it could be right then, which can be determined by grading prostate cancer cells based on their cellular characteristics. Trouble urinating, blood in the urine or sperm, bone pain, swelling of the legs, or pressure in the pelvic region are all signs of advanced prostate cancer. Men who are having these signs should get medical treatment right away. Unfortunately, symptoms are not always present, and this form of cancer can be ignored during regular tests. When cancer has spread to the bones, it is normal to feel extreme pain in the bones, which is a sure indication that the disease has spread. Prostate cancer must be able to spread across the majority of the prostate tissues until it can spread to other areas of the body. Some men have pain in their pelvis, neck, lower back, or even hip, and may be suffering with unexplained weight loss and exhaustion. The most frequent cause of urinary problems is a disease known as Benign Prostate Hyperplasia, or BPH; this refers to a swollen prostate, which induces intense discomfort in the urinary tract. BPH is not a precancerous or cancerous disease, but it is normal in men over the age of 50. Men must do their utmost to ensure a safe future free of prostate cancer; therefore, feeding and dieting correctly would be the best choice.

Fortunately, for every man suffering from prostate cancer, there are a number of medications available to assist in the management of this condition. Radiation therapy, surgery, and hormone manipulations are the four primary treatment choices for prostate cancer. Treatment methods differ based on the level of prostate cancer and which approach is more helpful to the patient. Surgery to remove the entire prostate is the most effective procedure for all patients in the early stages of prostate cancer. This is known as radical prostatectomy, because it involves the removal of the seminal vesicles, prostate, and vas deferens. In general, prostate cancer is graded before starting any therapy or resorting to any treatment strategy based on the visible characteristics of prostate cancer cells in one's body as well as the physical characteristics visible in an individual; this is where deep structures come into play as a treatment stage/mechanism for prostate cancer. The use of deep structures in prostate cancer grading comes from the fact that the

tool allows the interpretation of cells' characteristics based on cell characteristics as well as improvements in cell DNA structures, which is a frequent symptom of prostate cancer infection.

2.5 Diagnosis of Prostate Cancer

2.5.1 Screening

Prostate cancer can be diagnosed early by screening. Prostate cancer screening consists of two types of tests: a blood test for prostate specific antigen (PSA) and a digital rectal examination (DRE). PSA tests diagnose cancer by measuring the amount of PSA (a particular protein formed solely by the prostate) in the blood, while DRE assists physical analysis in identifying anomalies or prostate thickness.

Although PSA and DRE are very useful tests for early detection of prostate cancer, these two are not always effective and accurate. PSA test may show abnormal results ("false positive") when a person is in fact healthy or it may show normal results ("false negative") even if a person has cancer. Studies suggested that after the initial introduction of PSA testing, the lifetime risk of being diagnosed with cancer increased from one in eleven to one in six, whereas lifetime death expectancy remained same that is one in thirty four [17].

2.5.2 Prostate Biopsy

If the PSA test shows elevated PSA level, your doctor may recommend further tests to estimate the possibility of prostate cancer. Sometimes, PSA level may get high due to some other factors such as an infection or sexual stimulation. If the additional tests show less probability of prostate cancer, a person may avoid having a biopsy. Depending on test results and some other factors, such as family history, biopsy history and ethnicity, the decision to have a biopsy can be made. A biopsy is a minor operation that involves scraping small fragments of prostate tissue and testing them under a microscope for cancer cells [18]. This test is the best way to assess the occurrence of prostate cancer as well as the cancer grading, which shows how far the cancer has advanced.

2.5.3 Grading and Staging of Prostate Tumor

After the core biopsies indicate the presence of cancer cells in the tissue, the next step is to assess the cancer's level and grade. Grading is a mean of assessing how aggressive and unhealthy cancer cells are, and provides us with details about the cancer's seriousness. Staging refers to the scale and presence of a cancer, i.e. how large it is and if it has spread to other areas of the body. Both grading and staging are crucial measures in deciding how curable a cancer is and what treatment options are available.

2.5.3.1 Gleason Grading system

The keystone in cancer treatment is prostate carcinoma grading. In 1978, the American Cancer Society organized several seminars to finalize the prostate cancer classification system [19, 20]. The society compared different grading systems before accepting Donald Gleason's Gleason grading system, which he developed in 1966 [21] , to be the most powerful medium for cancer prediction, since it was discrete, simple, consistent and clinically relevant.

Gleason classification, which is based on analysis of Hematoxylin and Eosin (H and E) stained histological sections under the light microscope, is currently the most commonly applied method around the world. This ranking scheme is normal because it only recognizes the architectural patterns of prostatic tumors, known as Gleason patterns. Over the years, the Gleason ranking scheme has been revised many times. In 2014, the International Society of Urological Pathology (ISUP) conducted a systematic review and modification of this method, as well as the introduction of several important developments [20]. Each histological pattern is assigned a number ranging from 1 (well differentiated) to 5 (poorly/least differentiated) in the Gleason grading scheme. Gleason pattern 3 is characterized by irregularly spaced, well-formed glands of different sizes. Note that Gleason patterns 1 and 2 are no longer recommended because they show the same effects as grade 3. Fused glands, abnormal cribriform glands, glomeruli structures, and poorly-formed glands are all defined by Pattern 4. Comedonecrosis and poorly differentiated single cells, most likely forming cords and including vacuoles, are present in Gleason pattern 5.

The Gleason score is then calculated by multiplying the two most common trends in the specimen. The lowest score appointed is Gleason 6 (3+3), which proposes that the cancer is slow

growing. The highest score 10 suggests that the cancer is extremely destructive and high-risk. In the ISUP conference of 2014, a new grade grouping (from 1 to 5) [22] was suggested in order to make a more accurate cancer prognosis.

Gleason scores of 6 or less are assigned to grade category 1 in this scheme. Grade groups 2 and 3 are assigned to Gleason scores of 3+4=7 and 4+3=7, respectively. They were not previously known as GS 7 and were not labelled as such. Class 4 is made up of people with a Gleason score of 4+4=8, and groups 9 and 10 are made up of people with a Gleason score of 9 and 10 [23-26].

Figure 2.2 [23] shows the recent Gleason grading system with 5 group grades assigned to different Gleason scores each corresponding to a unique Gleason pattern.

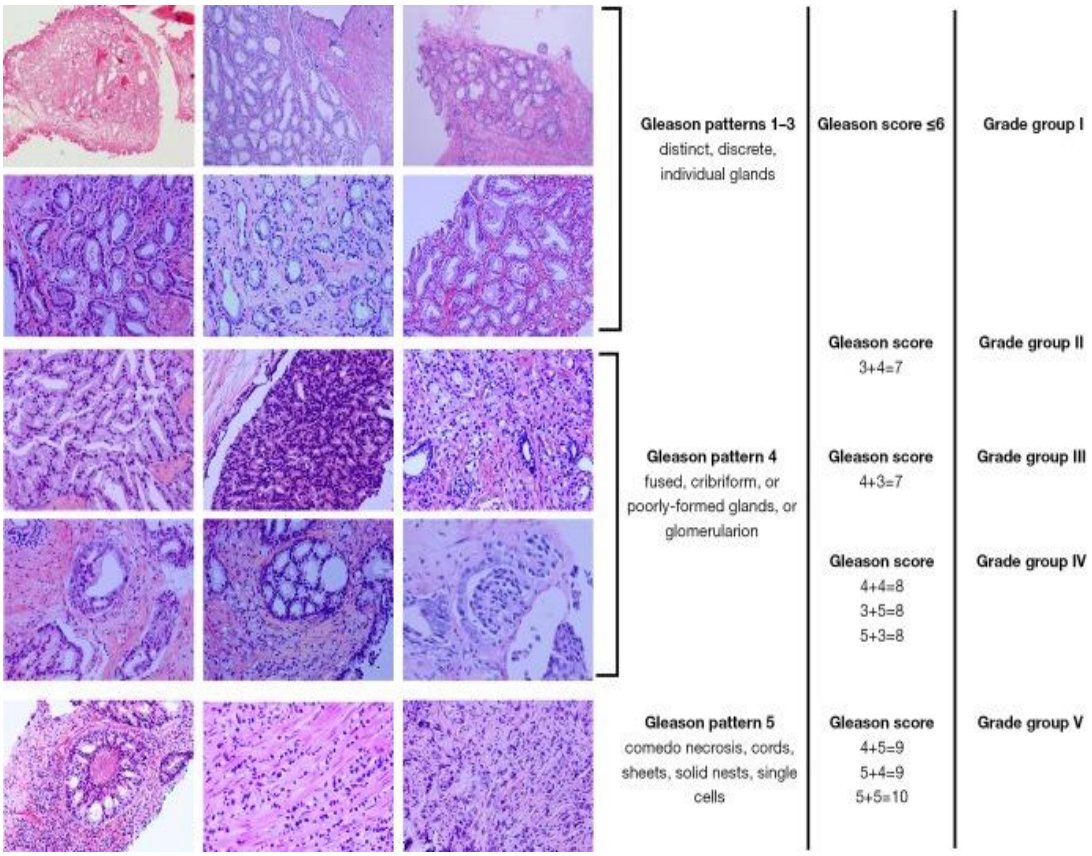

Figure 2.2: The contemporary Gleason Grading system having five group grades assigned to various Gleason scores[23]

Despite the fact that the Gleason rating scheme has improved over time, pathologists nevertheless administer the majority of grades manually. This is a time-consuming work and can only be done by highly educated pathologists, and still suffers from a lot of variation among pathologists' decisions [27, 28]. This issue is more common when distinguishing Gleason 3 from Gleason 4, which, if not distinguished correctly, may have a major effect on subsequent care [6-8]. An automated annotation method may be a viable solution to this problem. Feature-engineering methods for prostate cancer Gleason scoring were first used in previous studies (27-29) Finally, deep learning applications were used to diagnose cancer and grade biopsies using Gleason grading [29, 30].

2.5.3.2 Staging

Stage refers to the magnitude or extent of a tumor. Stage of cancer assists doctors to determine the seriousness of disease and recommend the best treatment. Sometimes a low stage cancer can be of high risk. The TNM scheme is used to categorize cancer levels, with T denoting tumor size and its spread to surrounding tissues, and N and M denoting nodes (tumor spread to adjacent lymph nodes) and metastasis (prostate cancer spread to other body parts), respectively. DRE and special imaging tests, such as ultrasound, MRI, CT, or bone scan, are used to assess the "T" level. The PSA degree, prostate biopsy, and DRE findings are used to determine cancer staging.

2.6 Prostate Cancer Screening and Technological Application History

Cancer screening has a high propensity for moral conflict because it subjects a large number of patients to risks and concerns in order to provide other people with the benefits of early detection. Prostate cancer is one of the most well-known male cancers. Despite the fact that routine PCa screening is still debatable, the American Cancer Society (ACS) recommends PSA monitoring for men at high risk. The use of PSA to test for malignant prostate growth (PSA) has greatly aided in the diagnosis and management of prostate disease. To detect any signs of prostate tumors, researchers used a prostate-specific antigen test in combination with a digital rectal inspection (DRE). Nonetheless, PSA has not often differentiated the precise degree of PSA to determine malignant prostate development, which has resulted in specialists treating immaterial tumors being confused.

Aversion of prostate cancer and a prime position are thought to be important factors in controlling inflammation and increasing patient resilience. [31] Recommend that, in order to prevent malignant prostate cancer, the importance of public health education be emphasized. According to patient promotion [31], initiatives based on the Health Belief Model had a significant impact on the resigned participants' prostate cancer prevention activities by enhancing their insight level and HBM components. The experts agree that by paying more attention to the instructional system and preparation based on instructional speculations and templates, an early diagnosis will be made and the illness can be successfully treated [31]. Furthermore, prostate cancer screening and management are more successful when patients are fully engaged in these programs. As mentioned, some of the elements [32], Individuals' dispositions and convictions, for example, have a significant effect on people's decisions about preventive social care and compliance with prescription medical regimens. According to analysts, social insurance professionals should continue to persuade men to seek out malignant growth screening. This will serve as a roadmap for designing educational manuals for analysts and human resources professionals, as well as influencing examinations that will increase interest in PCa early exploration.

Furthermore, joint decision-making between patients and physicians should be encouraged in order to achieve the patients' PSA screening goals. Patients must be able to combine and incorporate quantitative results in their decision-making as a condition of joint critical decision-making and hazard communication. The patient's ability to decipher probability and danger details is dependent on his or her numeracy of quantitative concepts. A possible question of low numeracy persists, which poses a threat to the joint decision-making process [33]. Elective strategies for communicating ideas of risk to patients, especially among those with lower numeracy and proficiency, may promote mutual critical leadership in the early detection of prostate disease.

Furthermore, evidence-based clinical data is a critical knowledge that will help patients make informed decisions about whether or not to perform PSA testing. Nonetheless, clinical data materials must be adequate, adjusted, and unbiased in order for patients to be informed of the substantial risk of over diagnosis [34]. Finally, becoming aware of the PSA exam can be an important consideration. Because of a lack of records, limited access to treatment, and a lack of education and encouragement regarding prostate cancer screenings, most men may suffer [34] Adult males would be willing to take on more influential responsibility and learn about their fitness

if healthcare providers have more assets and accessible wellness inclusion. In addition, with improvements in legislation affecting the delivery of medical care, malignant growth-related mortality could be predicted, and men's personal satisfaction could improve. These findings suggest that a long-term doctor-patient partnership should be upgraded to provide important information to adult male patients [35]. Tumor markers may be used to assess malignant progression, predict and screen management responses, and evaluate if the illness has returned following the treatment. In most cases, cancer markers cannot be used alone to detect cancer; they must be used in conjunction with other tests. Deep structures are being studied to see how they can be used in the early detection and screening of prostate cancer.

2.7 Digital Pathology and Technological Enhancement of Diagnosis

Pathology employs tissues and body fluid analysis to formulate correct diagnostic approaches for illnesses such as prostate cancer. Based on pathological approach of diagnosis, patients are first assessed for infection. When it has been ruled out that there is an infection, analysis is performed on the patient's cells to determine the level of prostate cancer infection, the stage of infection and many other factors about the disease. The analysis of these factors usually employ various approaches, an example is use of deep structures, to observe the patients' prostate cancer cells behavior and shapes. Technological application has greatly revolutionized the prostate cancer cells' grading. This has been achieved through the application of deep structures to identify variance in an individual's cells, for instance, identifying an infected and a healthy cell.

In other medical fields, technology has played a critical role in the transformation of pathology. It has been used in the infrastructure transformation in an area for lab-to-lab sharing of information by direct interfaces with the laboratory information and management systems. Technology is used in the provision of virtual pathology provisions in inaccessible areas and enables consultations of individual ideas. Moreover, pathologists apply technology in the integration of the laboratory test results management across multiple lab sites by use of differing lab information systems such as immunology, hematology, cytology, microbiology, and biochemistry. Consolidation by means of system integration can lead to enhanced accuracy in processing of data and facilitating the performance of multi-disciplinary groups. Pathologists can

share important information for enhancing patient care like enabling referrals and consultations and use of patient electronic health records. Technology enables patients, with prolonged conditions, to manage their health care by assisting them to self-manage their conditions. It has provided many other enhanced approaches that have made the diagnosis of various chronic disease relatively easier and even experimental.

2.8 Prostate Cancer Early Detection and its Contribution to Treatment According to US Healthcare System

Early detection and prevention following prostate cancer diagnosis have been made possible, thanks to the successful use of preventive medicine in the US healthcare system. Prostate cancer is the second most prevalent cancer in men in the United States, after skin cancer. It is more common in African American men than in Caucasian men. Men with a higher BMI (body mass index) are more likely to be diagnosed with prostate cancer and less likely to be diagnosed with other cancers [36]. Furthermore, race appears to be linked to an increased risk of prostate cancer. Prostate cancer is one of the most common diseases listed as "non-amenable." Prostate cancer was the second leading cause of cancer deaths for men in the United States in 2000, with 31,000 deaths. A lifestyle change, that emphasizes smoking avoidance, fitness, and weight stability in return reduces the risk of prostate cancer and facilitates safe living, is one of the key preventive methods for prostate cancer. Early detection or preventive drug therapy is an effective way to slow the development of prostate cancer by closely monitoring and stabilizing the infected region to prevent further spread. Detecting and curing prostate cancer before signs appear, according to the National Cancer Institute, does not increase your health or help you live longer. It is not easy for health care providers or doctors to monitor the quality of primary care programs for cancer patients or other amenable diseases. As a result, unique environmental conditions can also lead to early diagnosis of prostate cancer. According to the evidence so far, there is a significant positive relationship between good preventive medicine and early detection of diseases amenable to medical intervention in the US healthcare system. However, the crucial question is whether, if diagnosed early, action will reduce re-admission by using preventative drugs [16].

2.9 Background Related to CAD Systems

2.9.1 What is Computer Based Diagnosis System

Computer based diagnosis is a diagnostic technique that employs the use of multiple computer-controlled instruments to assist doctors in interpreting medical images based on their characteristics and in rendering the proper diagnosis of a condition so that the appropriate medication can be written. Picture recognition, image segmentation, image extraction and filtering, and classification are typically used in computer-aided diagnostics. Preprocessing aids are used in improving the image's accuracy and preparing it for subsequent processes such as segmentation, Feature Extraction. Segmentation is a technique for separating the suspicious mass field from the rest of the mammogram. The characteristics of a segmented mass area are extracted using feature extraction. The derived characteristics are analyzed, and the mammogram is graded as benign or cancerous depending on the results. Multi-parametric MRI is one of the most often used CAD techniques in the diagnosis of prostate cancer. The basic function of this method is to improve prostate cancer cell analysis in order to demonstrate the cell's embedded characteristics. One of the key reasons that has improved cancer cell grading ability in the use of CAD is improved cell structure specification. Computer-aided detection, CAD, has advanced the process of detection of infection patterns in the diagnosis of prostate cancer, MRI and other morphological characteristics. A method for distinguishing the suspicious mass area from the remainder of the mammogram is segmentation. Using function extraction, the features of a segmented mass area are retrieved. The mammogram is classified as benign or cancerous based on the outcomes of the derived characteristics. One of the most often used CAD methods in the detection of prostate cancer is multi-parametric MRI. The aim of this approach is to enhance prostate cancer cell analysis such that the cell's embedded features can be seen. Improved cell structure specification is one of the main reasons that CAD has improved cancer cell grading ability.

2.9.2 Application of Artificial Intelligence Algorithm for Prostate Cancer Diagnosis

Curing cancer is among the biggest challenges in the healthcare industry in 2021. While technologies and knowledge about cancer have increased immensely over the last two decades, the vast variability between different types of cancer and between patients who have the same cancer makes it hard to develop a definitive cure or detection mechanism. More than 100 cancer-related diseases require a comprehensive standard of care for managing these illnesses, including physical examination and continuous assessment and monitoring of patient's symptoms to keep up with the increased expanding nature of the disease. The challenges identifiable from this nature of cancer-related diseases in line with detection technologies include making cancer detection more accurate, high expense in the development of detection technologies, anatomic limitations, and issues with technologies for detecting MicroRNAs in cancer patients.

The significance of technological application in prostate cancer diagnosis is enhancement of the war against cancer, one of the most devastating disease conditions in the health landscape. This has been done through enhanced technologies for cancer detection to facilitate early detection of cancerous cells for effective treatment. These technologies will address and overcome flaws identified in the existing technologies, such as low specificity in cancer detection. The detection of cancerous cells in the human body is of paramount and integral importance in combating cancer and enhancing patient outcomes.

Technological application of prostate cancer treatment has been faced by multiple issues and has affected its enhancement considerably. Some of these issues include; low specificity. In the detection of any disease, the precision and efficiency of the available technologies and methods are crucial and non-negotiable. The detection of cancer cells using emerging technologies has to be highly precise to enhance early detection and effective treatment of various types of cancers before the cancer advance to lethal stages. However, the currently available cancer detection technologies have a low specificity and precision rate, making them unfit for use. In a study conducted to assess the specificity and sensitivity range of different technologies in the detection of cancer, the specificity of some technologies was found to be extremely low. The low specificity of cancer detection technologies found in this study was found in technologies such as MelaFind with a specificity range of 6-40%, Electrical Impedance Spectroscopy (EIS) technologies with a

specificity range of 34-55%, and Pigmented Lesion Assay (PLA) technologies 57-91%. With such low-range specificity, these technologies are unreliable in the detection of skin cancer and hence the need to increase the efficacy of these technologies. Lack of precision in cancer detection technologies is problematic, as the results projected from the screening process may be inaccurate. Another negatively impacting factor is anatomical limitation. Circulating Tumor Cells (CTCs) has been a significant area of focus as this liquid biopsy is integral in the early detection of cancer cells. Further CTCs are crucial in the evaluation of cancer recurrence and helps inform the best medical intervention as well as improve the efficacy of treatment procedures like chemotherapy. Nonetheless, CTC technology is limited by the harm it can cause to cancer patients, the bias in sampling, and difficulties in sampling deep tumors. In light of this, there is a need to improve these technologies to reduce the harm it can cause to patients as well as enhance the specificity of the technology in scanning and making a definite diagnosis of cancer in deep tumors. MicroRNA detection challenges in electrochemical, mechanical, electrical, and optical detection of cancerous cells also need improvement on precision, which is of utmost importance in modern medicine to enhance accurate diagnosis and treatment of cancer patients. And finally, due to high cost of technological integration, the cost of healthcare has been on the rise for the last few decades, which has increasingly reduced the accessibility to quality medical care. Both the government and international health organizations are always on the lookout for new technologies and strategies that can ease the disease burden on patients as well as prevent diseases from overwhelming the healthcare industry. Cancer treatment is one of the costliest healthcare encounters, and the costs are continually rising with the implementation of new technologies. High expense is the most significant hurdle against the massive implementation of cancer detection technologies. This is a problematic challenge that is barring people from accessing quality technology-based cancer screening and treatment. In this regard, more and more patients go for cancer screening when cancer has already advanced to lethal stages, making it hard for them to recover, particularly if the cancer cells have already spread throughout the body.

2.10 Prostate Cancer Statistics

Statistical data is important as it allows scientists to determine the need of additional research and where to focus their efforts. Understanding the international trends of prostate cancer would help health officials and scientists target the most needed area for better treatment, such as improving grading techniques and screening methods.

2.10.1 Global Cancer Ranking as a Cause of Death

Non-communicable diseases are currently the leading cause of death worldwide [37], and cancer is predicted to be the leading cause of global deaths and a significant impediment to longevity of countries across the world in the twenty-first century. Cancer is the leading cause of death for people under the age of 70 in 48 countries, according to WHO figures from 2015 (**Figure 2.3**), and it ranks second in 43 others. It is the third and fourth leading cause of death in an additional 22 nations [37].

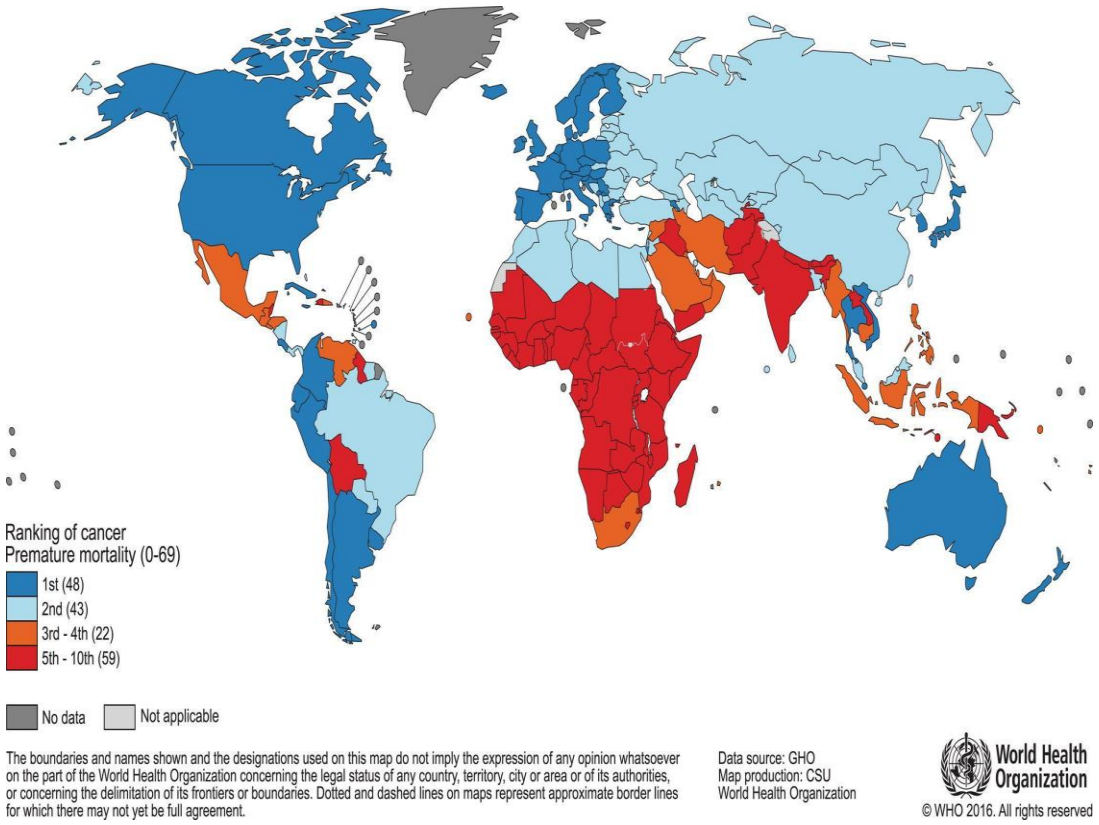

Figure 2.3: World map showing cancer ranking in 2015 as a cause of premature mortality (0-69) in 172 countries [37]

The growth of cancer burden across the world is mainly due to population growth and aging, and also because the prevailing risk factors for cancer, especially the ones associated with societal development and economic growth, have been changing over the recent years [38, 39]

2.10.2 Worldwide Cancer Patterns

Figure 2.4 and **Figure 2.5** show the cancer types which are most commonly found in different countries and responsible for the most mortalities. The maps show significant variation in cancer types at national levels. Prostate cancer is the most frequently diagnosed cancer in men in 105 countries, accounting for half of the overall 185, with Australia, Northern and Western Europe, the Americas, and the majority of Sub-Saharan Africa standing out. Prostate cancer is the leading cause of death among men in 46 countries, mostly in Africa, the Caribbean, and Sub-Saharan Africa.

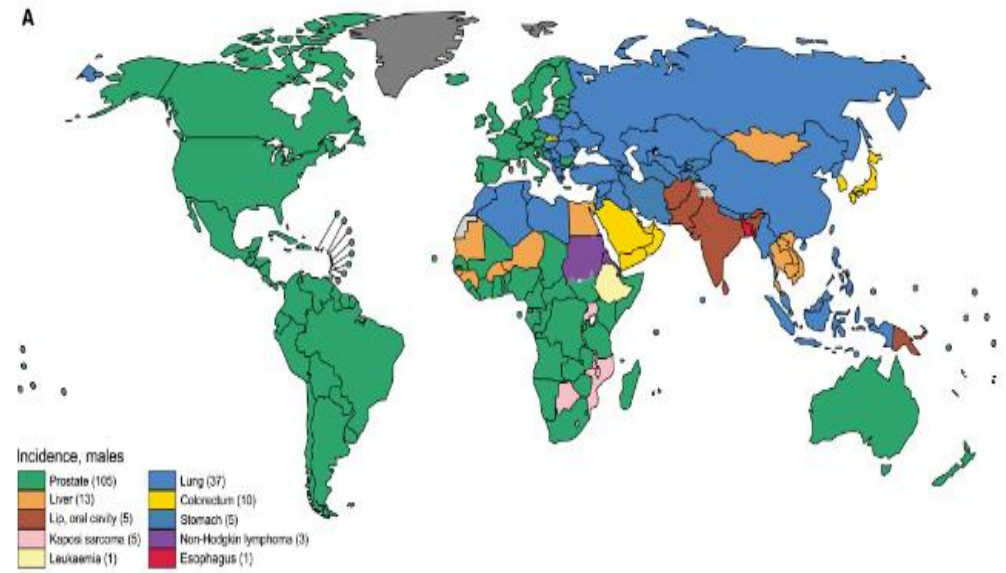

Figure 2.4: The most common cancer types in men in 185 countries. Source: IARC World Health Organization[39]

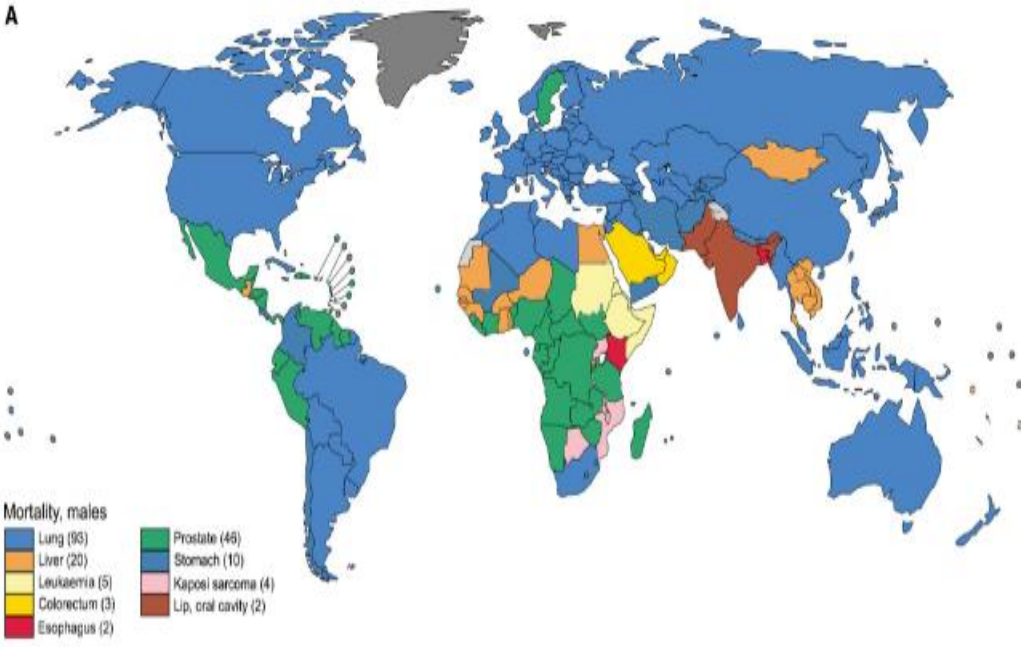

Figure 2.5: The cancer types which were considered as the major cause of deaths in 185 different countries. Source: IARC World Health Organization[39]

2.10.3 Incidence and Mortality Rates and Trends in Prostate Cancer

According to GLOBOCAN 2018 figures, nearly 1,276,000 new cases of prostate cancer are diagnosed worldwide in 2018, with 359,000 deaths due to prostate cancer. The more populated nations, such as Northern America, Western and Northern Europe, Oceania, and the Caribbean, have the highest occurrence rate [40] , whereas the lowest rates were noticed in Northern Africa, South-Central Asia and South-Eastern and Eastern Asia**.** On the other hand, the highest death rates were observed in Southern and Middle Africa and Caribbean, while Northern Africa and most parts of Asia remained least affected areas [41].

The given pie chart (**Figure 2.6**) represents PCa as the 2nd verily found cancer, constituting 13.5% of the total number of cases. 6.7 % of total deaths were caused due to prostate cancer, making it the 5th major reason for deaths in men.

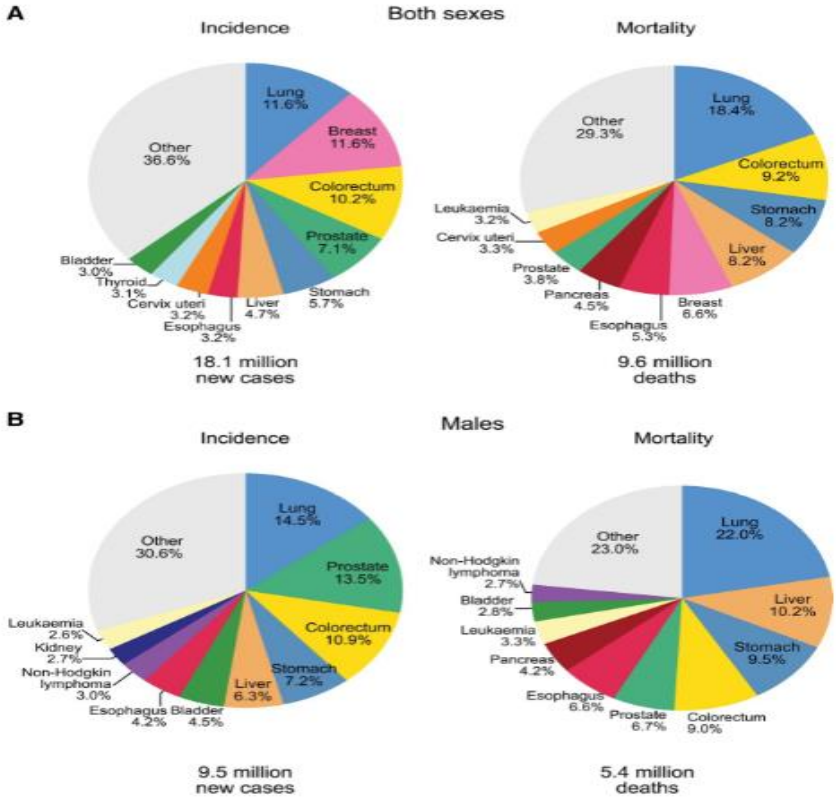

Figure 2.6: Distribution of Cases and Deaths for the 10 Most Maximum score Cancers in men in 2018[42].

Based on the GLOBOCAN database of 2018, age adjusted prostate cancer incidence and mortality rates during the most recent five years have been falling or showing a stable pattern in most of the countries across the world. The declining pattern is more prevalent in countries with human development. The decline in the use of PSA testing and better treatment methods might be the reasons for declined incidence and mortality trends respectively.

2.10.4 Key Statistics for Prostate Cancer in United States

Prostate cancer is fairly common in the U.S. compared to other cancers. It is the second leading cause of death in men in America, after lung cancer [43].

Estimated new cases of prostate cancer in 2020 in the U.S. were 191,930, which is about 10.6% of all new cancer cases. About 33,330 mortalities were estimated to occur in 2020, which makes up 5.5% of all cancer deaths [43]. From the above data, it is clear that prostate cancer is one of the most frequently occurring cancers across the globe, hence there is a need for an error-free system for grading and detection of prostate cancer. Manual grading system is a difficult task, which requires very expert pathologists, but still suffers from inter-and-intra observer variability. Moreover, every patient cannot have access to such expertise. Hence, a vigorous and consistent grading is needed at master level. In this thesis, we will develop a fully automatic deep learning system for grading of prostate biopsies.

3 Chapter 3: Literature Review

In this chapter, we'll go through the extensive feature engineering-based models that are used to diagnose prostate cancer automatically. Then, using convolutional neural networks, some recent deep learning-based models for segmentation and pixel level classification will be presented (CNN). Finally, we will review the Pixel level classification using UNET architecture upon which our proposed method is based.

3.1 Literature Review of Computer Aided Diagnosis (CAD) Systems and Digital Pathology

The goal of this literature review on prostate cancer grading using computer-based detection (CAD) in cancer TMA is to outline the steps involved in determining whether a patient has a malignancy or benignancy. In addition, the aim of this literature review is to look at the scientific advancements that have been made in the field of prostate cancer. Cancer is the most common cancer in women, followed by lung cancer and skin cancer, all of which are also common in men. This literature review was prompted by advances in prostate cancer screening technologies and the discovery of malignant cells. The recent advancements in tumor identification continue to draw attention to this field of research. The current study will look at the methods for image processing, segmentation, extraction and collection, and classification that contribute to the final declaration of the presence of malignant growths.

The method of detecting prostatic patterns in complex cancer TMA and other morphological characteristics has been improved significantly, thanks to computer-aided detection, or CAD. As a result, CAD systems have enabled the automation of many production and analysis tasks that would otherwise be performed manually by field professionals. The procedure aids in the accurate diagnosis and differentiation of benign and malignant growths, limiting or removing the need for needless biopsy. Although specificity varies, the technology has the highest sensitivity of all existing imaging modalities for prostate detection, ranging from 95.0 percent to 99.0 percent. During cancer tissue enhancement, an artificial CAD device produces color-coding based on signal amplitude, resulting in changes in voxels. It improves the pattern

analysis by allowing touch enhancement of the patterns through a range of photographs, allowing a more accurate diagnosis of prostate cancer.

Using image processing knowledge for diagnosis, literature shows researchers have utilized multiple methodologies to achieve different goals such as extracted features-based classification, image registration, and region and object segmentation. With the development in Computer Aided Diagnosis, pathological diagnosis has entered a new era. Now digital diagnosis is considered much precise and is used on large scale in diagnosis of prostate cancer and other cancer cell related infections.

3.2 Feature Engineering based Techniques' Review

Many traditional techniques of medical imaging focus on an image's pixel structure. Based on the corresponding pixel in the input image and its value in the function index, the value of the pixel in the output image is decided. The most popular morphological operation algorithm is area growth and is used to remove linked pixels from an image in the relative region. The parameters for similarities often depend on the set of values for pixel density or other characteristics. CAD Gleason grading for prostate cancer diagnosis has been the subject of several researches. The most popular technique is feature engineering based approach, in which features are derived from medical images, either whole slides or tissue micro array (TMA) images, and then fed into traditional machine learning classifiers like support vector machine (SVM) [44] , Bayesian classifier [45] or random forests [46].

Farooq et al [9] used Gabor filter and local binary patterns for features extraction. Gabor filter variants were created by rotating and expanding mother function. It has resemblance with the two properties of cells in the basic visual cortex, which are band pass nature and directionality. After assessing the potency of eight neighbors, the real version of LBP operator was carried on each pixel. Due to combination of multiple texture features, the suggested approach demonstrated higher accuracy. These selected features were used in KNN classifier to automatic grade the prostate cancerous cells. The KNN model achieved accuracy of 98.3% for real dataset of 268 histological E&H images collected from 160 different patients at different times.

The power distribution of histological tissue images was used by Smith et al. [47] to reflect texture characteristics of prostatic biopsies. They used nearest neighbor classifier to grade those characteristic features into Gleason grade 1, grade 3, grade 4 and grade 5.

Farjam et al. [48] suggested a multistage classifier based on morphometric and texture characteristics. Those morphometric and texture features are used to identify gland units. The image is then classified into grades 1 through 5 using morphometric and texture attributes derived from gland units in a sequence of classification levels. Since this strategy leverages properties such as roundness and shape distribution, which are associated with the structure of the glands and do not depend on magnification, the suggested approach is resistant to changes in lighting and magnification. It used two different datasets, one consists of 91 images with similar magnifications and illuminations and dataset 2 consist of 199 images with different magnifications and illuminations. It used T- structure algorithm to classify the Gleason grade into 5 corresponding groups. This algorithm achieved 95% accuracy on dataset 1 and 85% accuracy on dataset 2.

Nguyen et al. [49] used structural features of prostate glands to classify pre-extracted regions of interest (ROIs) into benign, G3, and G4. Based on structural features, used a hierarchical (binary) classification scheme which utilizes the two methods and obtains 85.6 % accuracy in classifying an input tissue pattern into one of the three groups. The data was used to make the classifier learn textural behavior and categorize each pixel into various classes. The segmented regions were then used to calculate the automatic diagnosis results. Quantitative information collected from glands, stroma was combined with Morphological features, and logistic regression was then used to distinguish between the regions of Gleason grade 3 and grade 4. The operating curve gave the accuracy of 82%, which is a perfect value when inter-observer variability is taken into account. The described papers achieved good results on their datasets due to high dependence on feature extraction.

In **[50] Jian Ren** used the structure features and Delaunay Triangulation. Based on these features, prostate glands are segmented into different regions. The uniqueness of this method lies in the fact that they carried out regional based nuclei division without using earlier knowledge of lumen. The respective glands can be categorized by employing Delaunay Triangulation and structure features since each gland region is neighbored by nuclei. The model comprised of three major steps.First, pigment variations are removed from various images through preprocessing.

Then, each nuclei as well as global gland regions are recognized. Afterwards, gland region from the distance map is grouped with nearby nuclei in order to construct distinct gland. Lastly, for each segmented gland a shape descriptor is specified. The precision, recall and F1 score achieved in this model is 0.94, 0.11, 0.60, 0.23 and 0.70, 0.19 respectively.

Nilgoon [51] used hyper spectral transmission images using sixteen light wavelengths, those images are converted into RGB images using Principle Component analysis. He proposed an effective technique for spotting glandular structures of prostate and its nuclei. They used hyper spectral transmission images using sixteen light wavelengths, which were converted into RGB images using Principle Component analysis. The segmented results were achieved by clustering method which was further cleaned by morphological methods. PCA is a method that reduces dimensionality of large data set by decreasing number of variables which still have most of the information. Since basis vectors are computed individually for each dataset, PC RGB composite images, having different colors for alike materials, are produced when PCA is run on dataset independently. For comparing RGB images, they transformed datasets into a single basis vector set and scaled equally so that colour intensities between RGB images may be examined. The PC RGB images were then converted into LAB color space. K-mean clustering was then applied by setting number of clusters equal to three. The gland class, which was impure and had some residuals of stoma, was then cleaned using morphological cleaning functions. It resulted in well segmented glands. Since the gland cells have cytoplasm and nuclei, the same procedure of K-mean clustering and morphological cleaning was carried out in order to obtain segmented nuclei. The segmented results are achieved by clustering method which is further cleaned by morphological methods. The accuracy achieved in this method is 80% of spectral images are correctly classified which is most of the 60% TMAs images.

In **[52] Angshuman** used morphological scale space for feature extraction and gland segmentation which is totally based on regional glands structure. The method's dependency was on a biological hint (that EL surrounds a gland which looks darker in H and E stained images) instead of gland distinct signatures of intensity or texture, which may differ across the whole slide. Hence, the difference in stain or gland structure over the slide didn't have any effect on this method. The performance of model is measured by calculating F1 Score, achieved 0.68 F1 score which consist of 85% of the histological images.

Table 3.1 LITERATURE REVIEW BASED ON FEATURE ENGINEERING TECHNIQUES

Author	Year	Dataset	Features	Classifier	Results
Farooq [9]	2017	Local Dataset	Texture Features	KNN	98.3 %
Farjam [48]	2005	Local Dataset	Morphometric and Texture Features	Multi-Stage	95%
Nguyen [53]	2017	Whole Slide Local Dataset	Structural Features	ROI based Binary Scheme	85.6 %
Jian [50]	2017	Local Dataset	Structural Features	Segmentation	0.90
Nilgoon [51]	2017	TMA (Vancouver Prostate Centre)	Morphological and Glandular Structures	Segmentation	80%
A Paul [52]	2016	GlaS Challenge dataset	Region based Structural Features	Segmentation	0.68

3.3 Deep Learning Techniques' Review

Deep learning techniques have shown favorable results in a variety of computer vision tasks, including segmentation, recognition, and object detection [54, 55]. These approaches use convolutional layers to create input images from which different characteristics can be extracted, ranging from low-level local features to high-level global features. A fully connected layer converts convoluted features into certain marks' probabilities at the end of the convolutional neural layers [54]. The batch normalization layer, for example, normalizes the input of a layer with a zero mean and a unit version [56]. A dropout layer, which is one of the regularization approaches that avoids randomly chosen nodes, has been shown to improve deep learning methods. To achieve persuasive performance, however, optimal layer combinations and structures, as well as precise fine-tuning of hyper-parameters, are needed [54, 55, 57].

Extensive research has been carried out based on deep learning methods to design automatic computer based model for accurately grading the cancer [58]. Deep learning based fully convolution neural network models are very useful in early detection of prostate cancer as compared to feature engineering based models [59]. Deep learning models are very successful in prostatic segmentation [60]. In early stage, CNN based architectures [61] are used for better feature extraction as compared to conventional feature engineering based methods. They analyze deep entropy features using different CNN models and pass those features to random forest classifier to predict the Gleason score. Augmented based technique is proposed in [62], which uses three different CNNs, combine their prediction results and predict the Gleason score by logistic regression method. They achieved 92% and 86% accuracy in classifying low and high prostate grade. [63] used both morphological and texture features achieving 79% accuracy of classifying benign with other higher grades.

A very recent studies [2] implement the prostate TMAs classification into four categories; benign, grade 3, grade 4 and grade 5. In this model CNN is used as directly for feature extraction and Mobile NetV2 architecture as backbone for classification into different prostate groups.

In **[64] Khani** used Gleason Challenge 2019 dataset which consist of TMAs (Tissue Microarrays). In this framework Gleason grading system is used to evaluate the prostate cancer. For achieving ground level annotation in MICCAI dataset, staple algorithm is used for training the model. The images patches are extracted with size 256x256 and 512x512 with an overlap of 8

pixels. Deep Labv3 with MobileNetV2 model is used as backbone for feature extraction as well as classification. Deep Labv3 model is consist of encoder-decoder based architecture which has been used in many semantic segmentation models. It achieved the quadratic Cohen's Kappa of 0.56.

Lokhande [65] proposed modified FCN8s with ResNet50 as backbone for classification on Gleason Challenge 2019 MICCAI dataset. This model is designed purely for semantic segmentation problem. The model implementation is done on MICCAAI 2019 dataset which consist of tissue microarrays (TMAs). The TMAs based annotation are used for pixel level classification, all the pixels have corresponding class according to pixel level. These pixel level annotations are than passed to model FCN8 with ResNet50 to classify each pixel of TMAs. The model is named is Carcino-Net and achieved average dice coefficient score of 0.74.

Bulten [66] used UNET model with different pixel spacing for the classification of prostate images into different grade groups. Due to limited dataset problem, on Harvard V1 dataset, data augmentation techniques are applied to achieve Cohen's kappa of 0.711. This model is totally encoder decoder based architecture. The system used randomly selected biopsies by the biopsy Gleason score from different patients. The Model is trained on 72 epochs with learning size of 0.0005 and batch size 8.

In **[67] Locus** used inception V3 CNN architecture as classifier on locally available dataset. He worked on Gleason grading score algorithm to grade the prostate biopsies. He introduced grade groups with different benign and prostate tissue. These grade groups based biopsies are than passed to CNN inception model and achieved average K of 0.70.

Shin used [68] self-attentive normalization (SAN) instead of normalization techniques on publically available Harvard dataset. To focus on more relevant parts of the feature map, the suggested approach can learn the element wise affine transformation. Because SAN only requires a few more learning parameters, it can be readily incorporated into current automated Gleason grading systems with minimal overhead. He used VGG16 CNN architecture for classification and achieved 79% accuracy as a whole.

Table 3.2 LITERATURE REVIEW BASED ON DEEP LEARNING TECHNIQUES

Author	**Year**	**Dataset**	**Classifier**	**Results**
Karimi [62]	2019	Local Dataset	Logistic Model	92%
Arvaniti [2]	2018	Harvard Dataset	CNN MobileNetV2	0.72
Khani [64]	2019	Gleason Challenge 2019	Deep Labv3+MobileNetV2	0.56
Lokhande [65]	2020	Gleason Challenge 2019	Carcino-Net	0.74
Bulten [66]	2019	Harvard Dataset	UNET	0.711
Lucas [67]	2019	Local Dataset	Inception V3	0.70
Shin [68]	2020	Harvard Dataset	VGG16	79%

4 Chapter 4: Experimental Methodology

We suggest an automatic computer-driven prostate cancer Gleason grading scheme based on deep learning and Convolutional Neural Networks in our study (CNN). Photos were taken from the Gleason 2019 Challenge dataset. Image selection, pre-processing, and deployment of various CNN architectures are the three stages in the suggested technique. The input images are then preprocessed and resize, which are then passed into a CNN architecture. The more detailed summary of these measures is given in subsequent headings. Our suggested model is depicted in **Figure 4.1** as a block diagram.

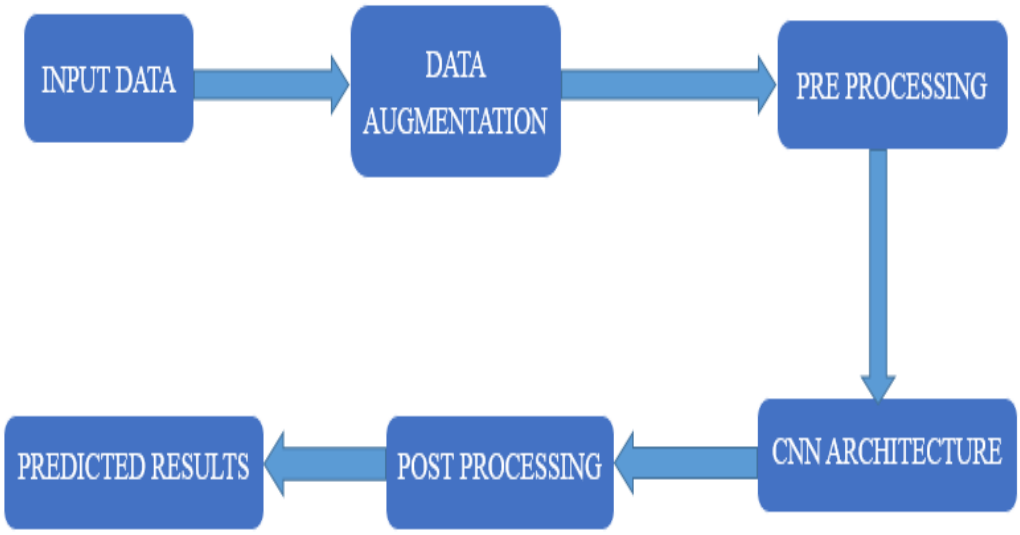

Figure 4.1: Diagram of Proposed Methodology

First, for understanding the dataset of prostate cancer images and tissue microarrays, local hospitals and doctors will be contacted. This will require a thorough understanding of the acquisition system and then labeling of the images by the pathologists. The dataset of images will be grouped and labeled into four categories; Grade 1 and 2, Grade 3, Grade 4 and Grade 5. Grade 1 and 2 are considered as healthy benign samples whereas Grade 3, 4, 5 are high cancerous sample images.

4.1 Data Preprocessing and Normalization

The Gleason 2019 Challenge dataset consists of tissue microarrays (TMAs) which are histological images obtained from patients who had radical prostatectomy. The TMAs were stained in H&E and scanned at 40x magnification with SCN400 slide scanner. These digitized TMA images were loaded into digital tab and stylish based android application which annotated the images under six experienced pathologists. These pathologists have experience of more than 20 years in the field of prostate cancer detection and grading. The below diagram shows the flow chart of preparing Tissue Microarrays (TMAs) and get them ready to feed in Convolutional Neural Network (CNN) after preprocessing.

The block diagram of our proposed scheme is shown in **Figure 4.1**. TMA images and their corresponding masks of Gleason Challenge dataset have resolution of 4608 x 5120. The TMA images have been resized to 512 x 512 for the 2 stages of training process. In PNG masks, the pixel values 0, 1, 2, 3, 4, 5 and 6 show corresponding Gleason score.

In Harvard Dataverse VI, the size of TMAs and masks are different, 3100 x 3100. We have resized the images to 512 x 512 for better generalization of model results which is trained on same size of images. In PNG masks, the pixel values 1, 2, 3 and 4 show corresponding Gleason score. After getting images, which are tissue microarray TMAs, data augmentation is applied to increase the data size.

4.2 Data Augmentation

Data augmentation is considered important, particularly in medical application fields where the amount of data is small. The overall performance of deep learning algorithm is improved using data augmentation. Data augmentation is the process of creating new training data based on existing training data by using methods like horizontal, vertical shift, horizontal flip, vertical flip, rotation, scaling and zooming. For data augmentation, we have used augmentor library [69], where tissue micro array images (TMA) are rotated from 90 to 270 degree with 0.25 probability and also performed left to right flipping. similarly top to bottom flipping is done with 0.15 probability. Finally we used random cropping operation with 0.35 probability. During data augmentation, we selected equal class ratio for augmentation to combat problems of data limitation. **Table 4.1** shows the parameters for data augmentation technique which we have used in the thesis.

Table 4.1: SUMMARY OF DATA AUGMENTATION IMPLEMENTATION

Methods	**Range**
Rotation	90 , 270 Degree
Flip Left to Right	15 % Probability
Flip Top to Bottom	25 % Probability
Random Cropping	40 % Probability

4.3 Convolutional Neural Network (CNN)

For automatic computer based model, to accurately assign the prostate grade and initiate the medication, CNN plays vital role due to its robustness and automatic feature learning capability. Before feeding into CNN, Images are pre-processed at certain levels and under computation resources for better results. Previous research has shown that deep learning approaches based on CNN can achieve high classification precision. This is due to the fact that deep learning methods are not restricted to handcrafted features, which can lead to a biased current domain information. With enough testing data and training, CNN models are able to get a clearer visual pattern, resulting in an increased precision. However, due to inter-variations in labelled results, training of these models is extremely difficult. While deep learning models have a strong reputation in the medical domain and computer vision, most medical applications have a small amount of branded training data. Therefore, programming a deep learning model with a vast number of parameters becomes very difficult.

CNN training requires large amount of data with considerable variation, which is often not available in medical diagnosis problems. This is solved by increasing the data using different techniques, named as data augmentation techniques. In our work, since the dataset is not enough to train the CNN model, we also applied data augmentation to solve the limited data problems and achieve state of art performance.

4.3.1 Convolutional Neural Network (CNN) Architecture Design

The input layer, hidden layers, and output layer make up a Convolutional Neural Network. Convolution layer, pooling layer, activation layer, normalization layer, and dropout or totally linked layer are all examples of hidden layers. Each of these layers has its own functionality which distinguishes the architecture from other conventional machine learning models. However CNN must have input, convolution, activation and fully connected layers. These layers lie on top of each other and make a hierarchical fashioned feature map. When it reaches the final layer, each layer takes data from the previous layer and transfers these features to the next layer. Softmax is normally preferred as the final layer. Based on the input function maps, the Softmax layer generates a probabilistic distribution of data. The Convolutional Neural Network's building block is the Convolution layer.

A convolutional layer performs convolution with different size of filters leading to the generation of feature maps. The size of the filter can be 3x3, 5x5, 7x7 etc. The shapes of convolutional filters are usually like boxes, as one of the images present in the data, and used to extract information from it. The resultant feature map consists of very useful and contextual information which is used in training. These feature maps correspond to some hidden layers which are controlled by activation layers.

There are various kinds of activation layers used on convolutional neutral network such as ReLU, TANH, Sigmoid and Softmax. The proposed network consists of 7 or 4 inputs depending on the size of TMA images. The pixel values range from 0 to 6. The input is assumed as one input with class labels which are act channels. Each channel is represented as red, blue, green channels of color images. **Figure 4.2** shows a simplest CNN architecture design, containing convolution, pooling and fully connected layers.

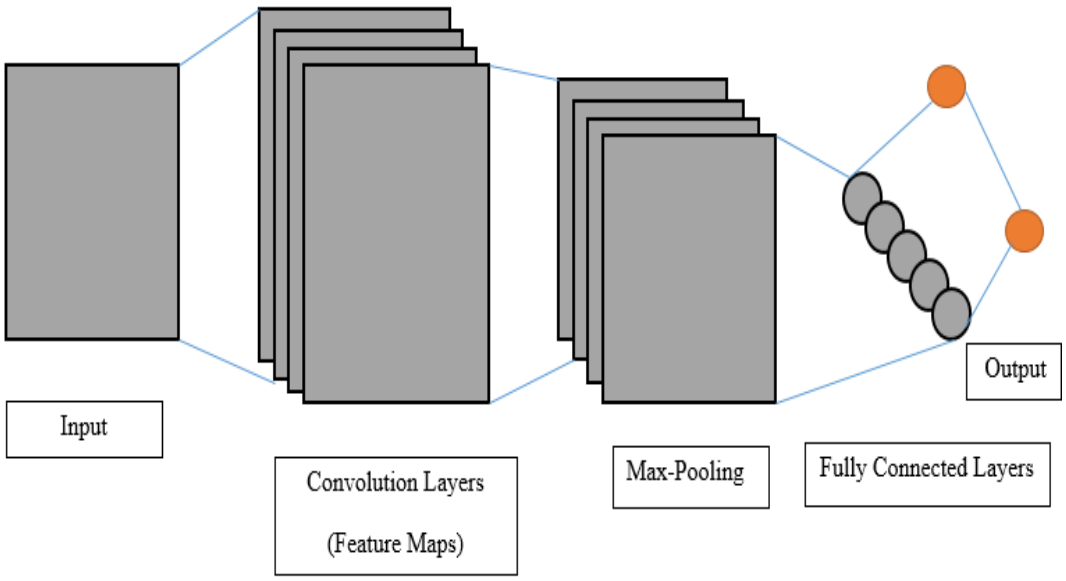

Figure 4.2: Simple Convolutional Neural Network which consists of input, convolutional, pooling, fully connected layers and output layers.

The Convolutional Kernels and connected weights are updated by a method called back propagation algorithm [44]. In Back-Propagation, input images are passed through the network via

feed forward pass and the network makes some class prediction. These predictions are then compared with the actual prediction. The resultant error in prediction is forwarded starting from the last layer to the first layer of the network.

4.4 UNET Architecture.

UNET architecture, as shown in **Figure 4.3**, was first introduced by [59] in 2015 for biomedical image segmentation. This model makes its place in the field of medical image segmentation in recent times due to its uniqueness. UNET model has the capability of contracting the input images into multiple feature maps. After contraction, it uses its previous feature maps to expand till it reaches its output level.

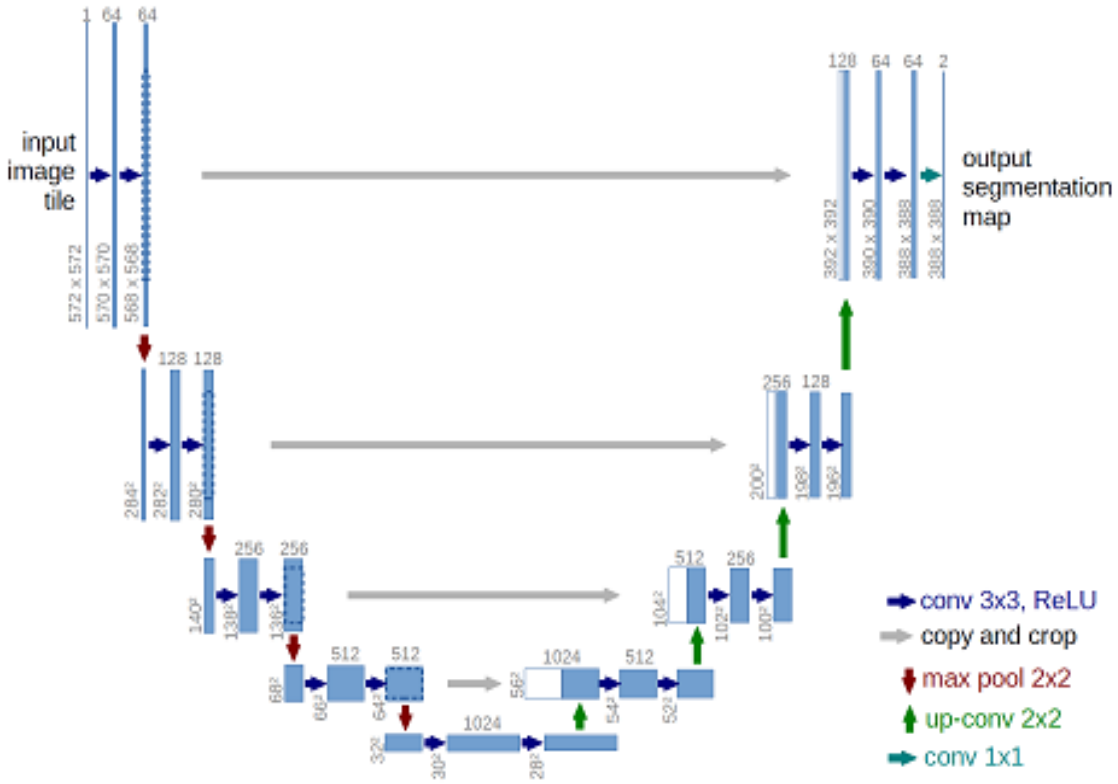

Figure 4.3: UNET Architecture with each stage shows encoding and decoding. Each blue box represents the feature maps followed by convolution and max pool [59].

Due to its contraction and expansion capability with concatenation power, it preserves the structural integrity of images. The layer details of UNET model with all four CNN architectures are given in **Table 4.2**

Table 4.2: UNET LAYERS OF ENCODING OF OUR FOUR ARCHITECURES

Model	Input Shape	Output Shape	Convolution layers	Pooling Layers
VGG19	512x512x1	512x512x6	16	5 max pool
ResNext50	512x512x1	512x512x6	48	1 max pool, 1 global avg pool
MobileNetv2	512x512x1	512x512x6	10	1 avg pool
ResNet50	512x512x1	512x512x6	48	1 max pool, 1 avg pool

4.5 System Architecture Setup

Deep learning based models require large amount of data for their training to perform well on test data. So data augmentation is applied using Augmenter library to reduce the over fitting problem and increase the generalizability of model. After properly resizing the TMAs images and data augmentation, TMAs are fed into convolutional neural network (CNN). Next step is to extract feature maps which is very critical for pixel level classification, because bad features may lead to poor pixel level classification results. Convolutional Neural Network is used for direct feature learning from data. In available datasets, large amount of class imbalance exists due to which model shows over fitting. To reduce class unequal problem, we used two phase training which solves the class imbalance problem. In first phase of training, we used equal class ratio to train the model and then used those weights to train on true class ratio. This methodology leads to achieve good results.

We used four different CNN architectures with UNET. The four architectures VGG19, ResNet50, Mobilenetv2 and ResNext50 are used for extracting progressive features from pre-processed TMAs. UNET model used these progressive features, up sample them, concatenate the features and generate predicted mask same as original mask. The mask contains the pixel level classes. Due to ResNet50 residual property and greater number of trainable parameters on both

MICCAI and Harvard Dataverse V1 datasets, we have achieved state of the art results. The details of architectures' parameters are given in **Table 4.3** and **Table 4.4**.

Table 4.3: DETAIL OF PARAMETERS OF FOUR ENCODER ARCHITECURE WITH UNET ON MICAAI DATASET

Model	Trainable	Non Trainable	Total
VGG19	29,058,807	4,032	29,062,839
ResNext50	31,993,850	70,214	32,064,064
MobileNetv2	8,012,215	36,096	8,048,311
ResNet50	32,514,426	47,558	32,561,984

Table 4.4: DETAIL OF PARAMETERS OF FOUR ENCODER ARCHITECURE WITH UNET ON HARVARD DATASET

Model	Trainable	Non Trainable	Total
VGG19	9,033,988	20,028,416	29,062,404
ResNext50	25,121,345	7,140,105	32,261,450
MobileNetv2	5,822,020	2,225,856	8,047,876
ResNet50	9,059,079	23,502,470	32,561,549

4.5.1 ResNet50.

In this architecture design, we have used ResNet50 [55] as an encoder part of UNET with skip connections and double convolution layers which are stacked parallel to each other.

ResNet is a deep convolutional neural network that is combined with images and auto-encodes. Resnet50 architecture consists of blocks which are convolution layers at different stages followed by pooling layers. Each block of encoder ResNet50 ends with global average pooling layer. Our residual network is made up of 177 layers and was derived from a 50-layer residual network architecture, as seen below in **Figure 4.4** . Each convolution block is followed by max pooling operation. At the end of convolution, average pooling layer is used.

We used only convolution layers block for segmentation of Prostate cancer. During convolution and pooling, the contextual information of tissue microarray (TMA) images is retained and then decoded with concatenation. ResNet50 performance is greater than all other three encoder architectures. Due to identity mapping property, gradient loss problem is solved and network learning speed increases. All the implantation is dependent on our system setup and hardware setup and this model gives best results on it.

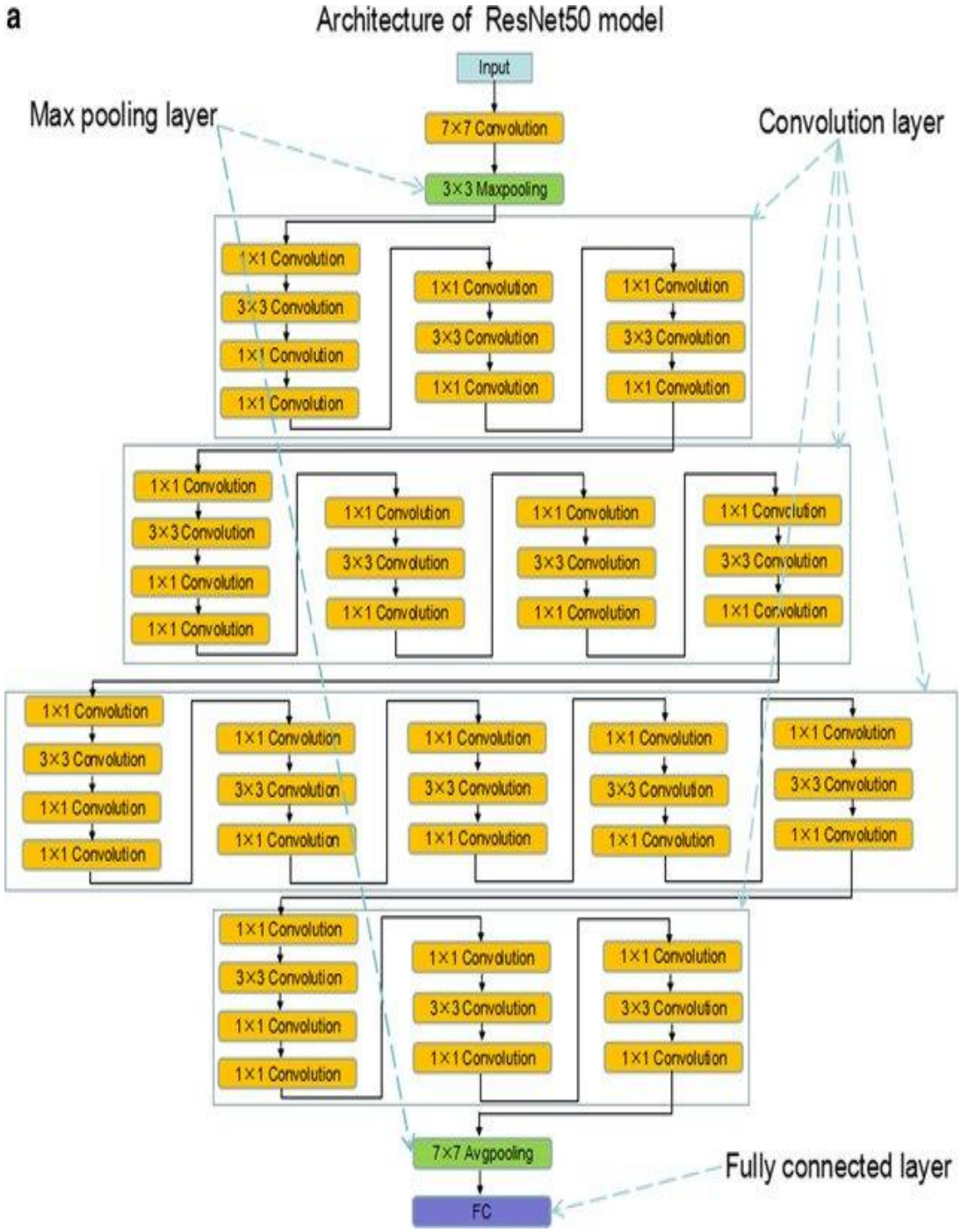

Figure 4.4: The architecture of ResNet50 and deep learning model flowchart [70]

4.5.2 VGG 19 Architecture

VGG stands for Visual Geometry Group which is one of the important deep learning architecture used for segmentation and classification. There are two types of VGG architecture in study in the literature named as VGG-16 and VGG-19. VGG-19 is the expansion of VGG-16, the only difference between the two architectures is that VGG-16 consists of 12 convolutional layers while VGG-19 consists of 16 convolutional layers. In VGG-19 (as shown in **Figure 4.5**), each convolutional layers block is followed by max pooling and at the end of architecture, three FC layers are followed by softmax activation function. The last layer of VGG-19 has 1000 neurons which are equal to the number of categories in ImageNet dataset. We used only convolutional layers block followed by max pooling for segmentation of prostate cancer.

Figure 4.5: VGG-19 architecture [71]

The default input size is 512x512x3 as used in VGG-16. The input is moved to a stack of convolutional blocks with most of the convolutional layers having a window size of 3x3 and all of the convolutional blocks being accompanied by the max pooling operation. In the case of all convolutional layers, the stride and padding are set to 1. Max pooling is performed with a 2x2 window with a stride of 2. ReLU is used as activation function in all the convolutional layers. The total number of parameters is 29,062,839 with input size kept as 512x512.

4.5.3 ResNext50 Architecture

We used ResNext50 architecture shown in **Figure 4.6** as an encoder to UNET for our automatic segmentation of prostate cancer. ResNext50 consists of convolution block followed by cross connections in each of its residual block. The map size for each feature map starts from 160 and becomes half in each operation i.e160, 80, 40, 20, 10. Up-sampling is then used to reconstruct the original map. **Figure 4.7** shows the complete structure of ResNeXt50 architecture with UNET.

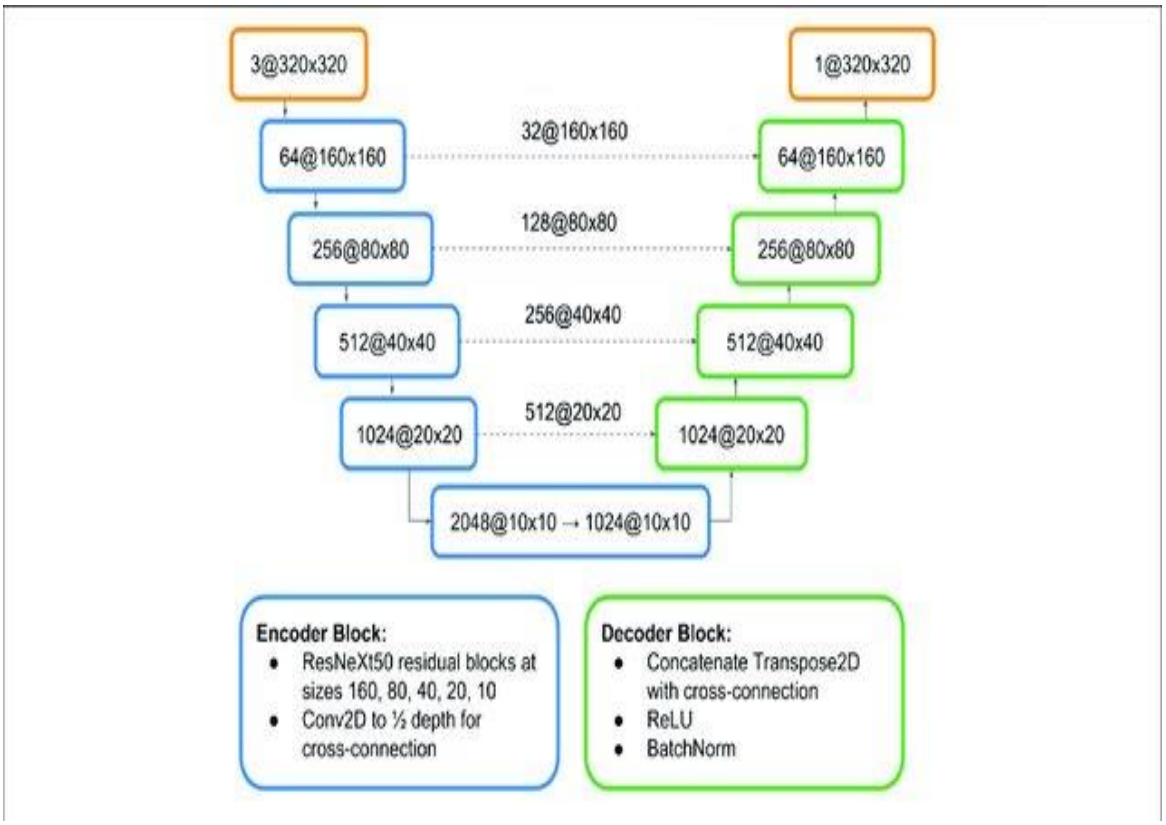

Figure 4.6: Architecture sketch for the U-Net inspired model using a ResNeXt50 encoder [72]

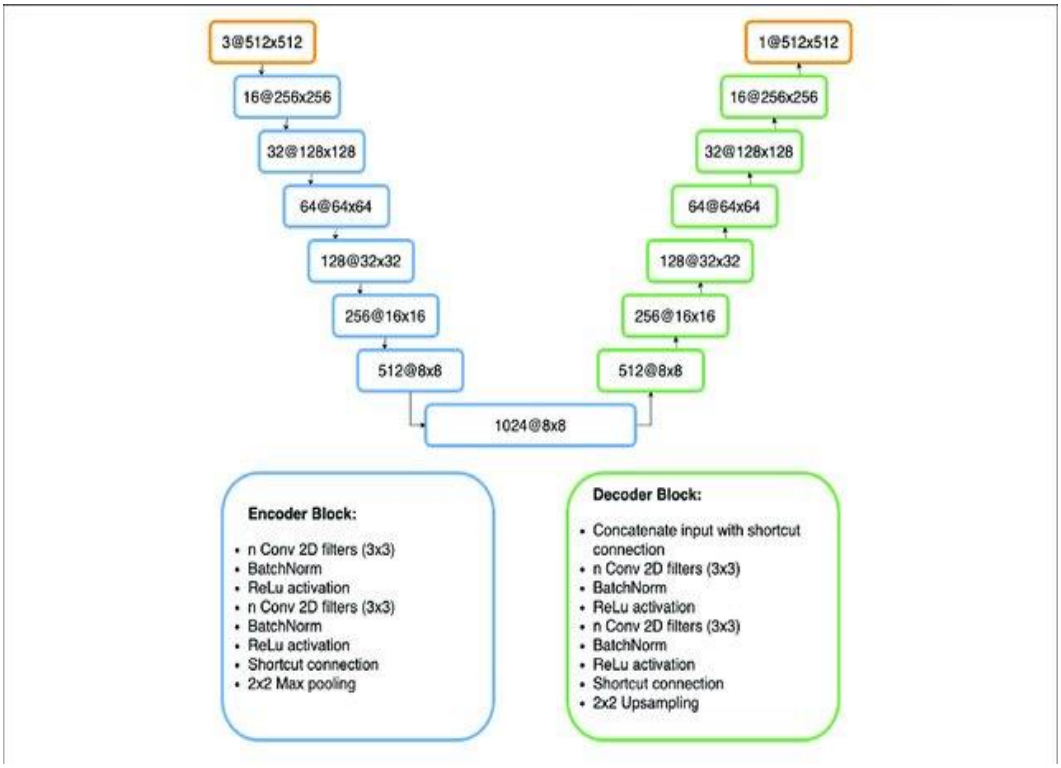

Figure 4.7: Architecture of the Deeper U-Net mode [72]

4.5.4 Mobile NET V2 Architecture

MobileNetv2 architecture is one of the famous architecture used in deep learning segmentation and classification. It is used in our segmentation methodology as an encoder to UNET due to its low value of inference time and high accuracy. The bottleneck residual blocks consist of 19 original fundamental blocks in the MobileNetV2 network architecture. Each block is followed by 1x1 convolutional layer with an average max pooling operation. The detail structure of MobileNetv2 Architecture is shown in **Figure 4.8**

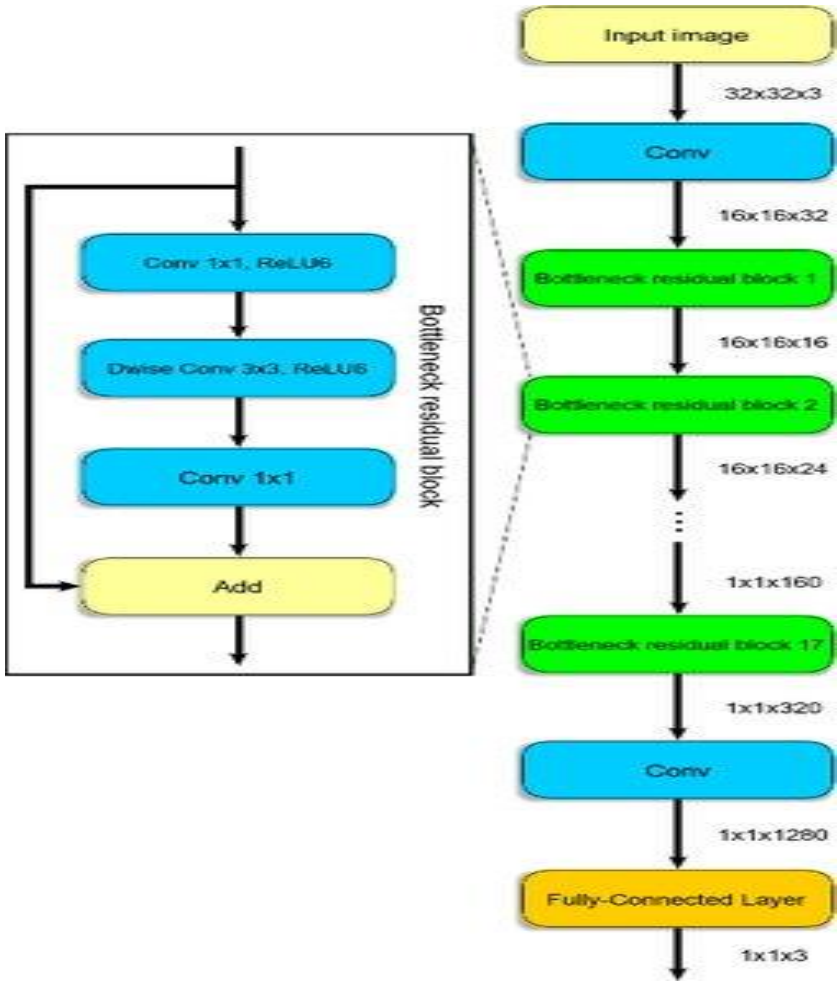

Figure 4.8: The architecture of the MobileNetv2network [73]

4.6 Loss Functions

In the development and enhancement of deep neural networks, the loss function is important. The loss function is a crucial component of both the Neural Network and the Convolutional Neural Network. They are characterized as a criterion against which the performance of the network is measured. Since our model optimization is entirely dependent on pixel level classification, we chose the mean square loss function.

We experimented with various loss functions during training to fine-tune our model, and our aim is to obtain state-of-the-art results on Gleason on both datasets. The loss functions that we used are as follows:

4.6.1 Stochastic Gradient Descent (SGD)

We used SGD as an optimizer at the outset of our training because it minimizes the calculations enormously. SGD takes one data point at random from the whole dataset at each iteration. The term ‘stochastic' refers to a process that is based on chance. As a result, instead of collecting all of the data for each iteration, Stochastic Gradient Descent selects a few samples at random. Gradient Descent has a definition called "batch," which reflects the total number of samples from a dataset used to calculate the gradient for each iteration.

4.6.2 Adam

Adam can be known of as a crossover between RMSprop and Stochastic Gradient Descent when it comes to momentum. It scales the learning rate using squared gradients, similar to RMSprop and it uses momentum by using the gradient's moving average rather than the gradient itself, similar to SGD with momentum. Adam is an adaptive learning rate process, which means he learns at his own pace.

4.7 Evaluation Metrics.

To assess the efficiency of segmentation models, evaluation criteria are used. We have used standard evaluation metrics that are currently being used in automatic Gleason grading of TMA images and also being used in clinical process of calculating inter observer variation. Our literature reveals that TMA based datasets are evaluated on Cohen’s Kappa [74] , F1 Score and Dice score [75]. We have also used these metrics to assess the performance of our proposed model. Another reason for choosing these metrics is to perform comparison with previous reported results in literature on automatic Gleason grading.

$$\text{Cohens Kappa} = \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}} \qquad (4.1)$$

Equation (1) shows the Cohen's Kappa where Po is observed agreement among ratters and Pe is hypothetical probability of chance agreement.

$$\text{F1 Score} = 2\left[\frac{\text{Precision}\,.\ \text{Recall}}{\text{Precision} + \text{Recall}}\right] \qquad (4.2)$$

F1 score is the function of Precision and Recall. The calculation of precision and recall is dependent on true positive and the sum of true positive and false positive which is shown in equation 3 and 4.

$$\text{F1 Score} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \qquad (4.3)$$

$$\text{Dice Score} = \frac{2 \times \text{TP}}{(\text{TP} + \text{FP}) + (\text{TP} + \text{FN})} \qquad (4.4)$$

$$\text{Overall Score} = \frac{\text{Cohens Kappa} + \text{F1 Score}}{2} \qquad (4.5)$$

We calculated the overall score consisting of both Cohen's kappa and F1 score for evaluating our model.

5 Chapter 5: Experimental Results

In this section we will briefly discuss the experimental results. We have performed our experimentation on Gleason Challenge 2019 and Harvard dataset. Both datasets contain tissue microarrays (TMAs) which have pixel level annotations by expert pathologists. Gleason Challenge 2019 dataset contains 244 TMAs, among which 188 are used as training by applying data augmentation, 33 TMAs are used as independent validation and 23 are used as testing cohort. While on Harvard dataset we have 509 TMAs as training, 133 TMAs as validation set and 245 TMAs as independent test cohort. Both datasets contain Gleason score of 3, 4 and 5.

5.1 Datasets

5.1.1 Gleason Challenge 2019 Dataset

The recently published dataset from the Gleason 2019 challenge has been used in our work [5]. Tissue microarray (TMA) images are included in this competition. Several specialist pathologists with years of experience in their fields annotated each TMA picture in great detail. Data is prepared by pathologists on the basis of majority voting annotations. The data contains samples belonging to benign and 3 different grades; G3, G4 and G5. The distribution of dataset into training, validation and testing cohorts is given in **Table 5.1**.

Table 5.1: THE DISTRIBUTION OF GLEASON GRADE IN THE TRAINING, VALIDATION AND TEST COHORTS.

	Total Cases	**Benign**	**G 3**	**G 4**	**G 5**
Train	188	72	111	134	10
Validation	33	13	20	23	1
Test	23	15	10	14	3
Total	244	100	141	171	14

Training images contain 188 TMA of prostatic tissues with benign, Gleason 3, Gleason 4, and Gleason 5. For testing the deep learning model we used 23 TMAs and for validation we used 33 TMAs. Some of the images from MICCAI dataset have been shown in **Figure 5.1**. Grade 3 and Grade 4 have high class ratio of images in data which create over fitting problems which are addressed by batch normalization, dropout and data augmentation.

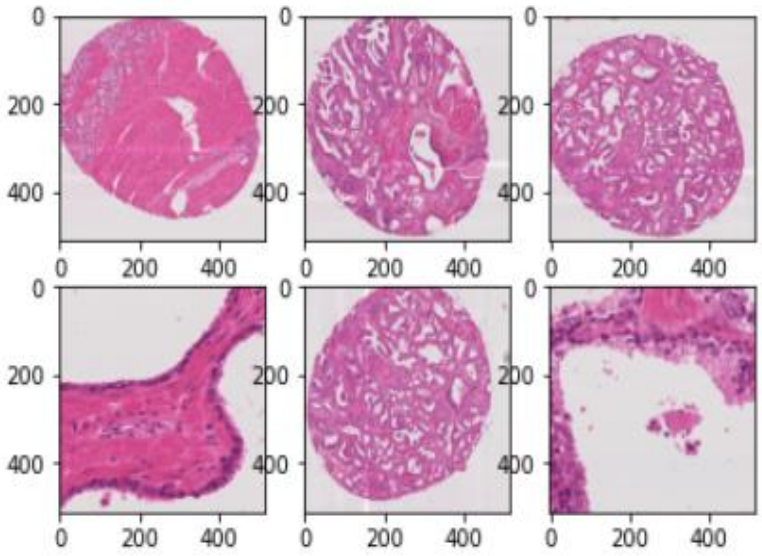

Figure 5.1: Example of Tissue Microarray Images (TMA) of MICCAI Dataset

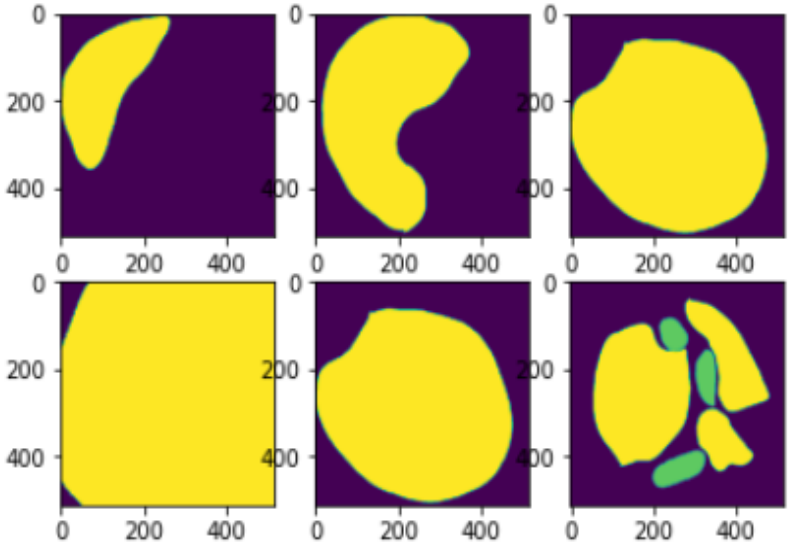

Figure 5.2: Example of true masks of corresponding TMAs.

Above TMAs are achieved after preprocessing of original images with corresponding masks after data augmentation. **Figure 5.2** shows ground truth of original TMAs which consist of Gleason score ranging from 0 to7.

5.1.2 Harvard Dataverse V1 Dataset

Harvard dataset is acquired from online databases of Harvard [2]. It contains five tissue micro arrays TMA each with 200 to 300 spots, which makes up the data. Objects or non-prostate tissue with spots (e.g. lymph node metastasis) is excluded from the study. The first pathologist (K.S.F.) identified the prostate TMA spots by carefully delineating cancerous areas and giving each one a Gleason pattern of 3, 4 or 5. TMA spots without cancerous areas have also been discovered to be benign. The distribution of Gleason scores across different tissue microarrays is seen in **Table 5.2**. Since TMA 80 has the most events, it has been assigned as the study cohort. TMA 76 was used as a confirmation cohort because it has the most evenly distributed Gleason ratings. Three other TMAs were used as a training cohort as a result. A second pathologist annotated the TMA spots in the research data separately, allowing the inter-pathologist variability to be quantified.

Table 5.2: THE DISTRIBUTION OF GLEASON GRADE IN THE TRAINING, VALIDATIO AND TEST COHORTS.

	Benign	**G 6**	**G 7**	**G 8**	**G 9**	**G 10**	**Total Images**
TMA 76	42	35	25	15	2	14	133
TMA 80	12	88	38	91	3	13	245
TMA 111	0	95	39	69	16	8	227
TMA 204	0	1	17	25	8	69	105
TMA 199	61	69	2	26	2	1	176

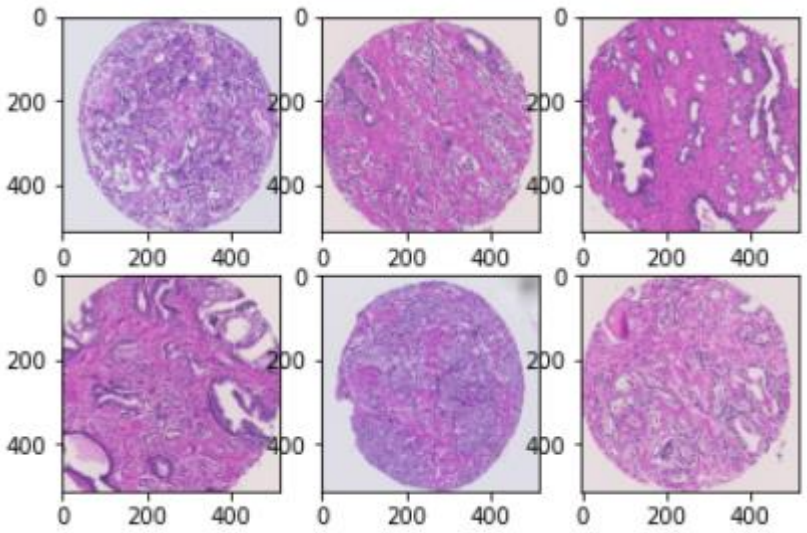

Figure 5.3: Example of Tissue Microarrays (TMA) of Harvard Dataverse V1 Dataset

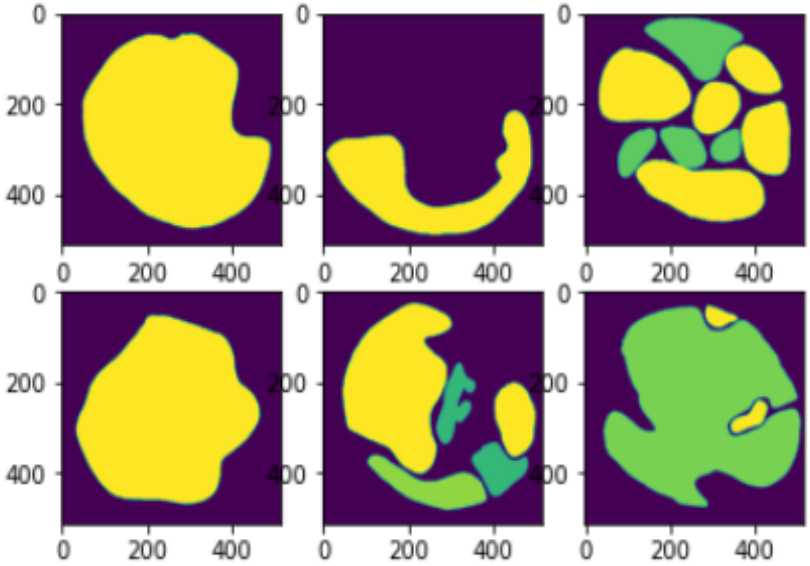

Figure 5.4: Example of True masks of correspond

Figure 5.3 and **Figure 5.4** represent the tissue microarrays TMAs and their corresponding masks of Harvard Dataverse V1 Dataset [2]. The above TMAs are achieved after preprocessing of

original images with corresponding masks after data augmentation. **Figure 5.4** shows ground truth of original TMAs which consist of Gleason score ranging from 0 to 5

5.2 Results on Gleason Challenge MACCAI 2019 Dataset

For our experimentation on automated grading of prostate cancer tissue using deep learning techniques [46, 62], we used the dataset from the MACCAI 2019 challenge [46, 60], which was part of the MICCAI 2019 meeting. Tissue microarray (TMA) images are annotated in depth by many specialist pathologists in this dataset. We have split the dataset of 244 TMA images into preparation, testing, and validation images. The dataset division is listed above.

In MICAAI dataset, four different CNN architectures are used as our UNET model encoder. We have achieved good results on all four encoder architectures and evaluated those using Dice Score, Cohen's kappa and F score. ResNet50 has performed well as compared to other encoder architectures due to its residual property and faster convergence. Categorical cross entropy is used as loss function with learning rate of 0.0001. ResNet50 has achieved overall score of 0.728 which is highest as compared to other models.

5.2.1 Comparison of Results on MICCAI Dataset

Table 5.3: COMPARISON OF UNET BASED MODEL RESULTS ON MACCAI DATASET

UNET Model	Encoder Backbone	Dice Score	Cohen's Kappa	F Score	Overall Score
1	VGG-19	0.49	0.30	0.31	0.305
2	ResNext 50	0.47	0.28	0.29	0.281
3	MobileNET	0.63	0.63	0.61	0.621
4	**ResNet 50**	**0.68**	**0.72**	**0.73**	**0.728**

Results in **Table 5.3** clearly indicate that when ResNet 50 is used as backbone with UNET, it outperforms all other architectures. The dice score for ResNet50 exceeds MobileNetv2 by 7.5%. While the increase is more significant when compared with ResNext50 and VGG-19 scores. This score go past ResNext50 and VGG-19 scores by 31% and 28% respectively. Similarly, ResNet 50 shows better results than other frameworks for both Cohen's Kappa and F score. A significant rise of 0.42 and 0.44 can be seen in overall score of our encoder from VGG-19 and ResNext50. However, there is an increase of 14.7% from MobileNetv2, proving it to be the best encoder for Gleason score assignment. Due to its residual blocks and identity mapping, ResNet50 has produced optimal feature maps. Those optimal feature maps contain all the pertinent features which can perfectly classify the image to its ground-truth class that is why ResNet50 gives state of art results as compared to other encoder architectures.

Table 5.4: COMPARISON OF OUR MODEL WITH BEST PERFORMING MODEL IN TERMS OF COHEN'S KAPPA ON MACCAI 2019 DATASET.

Model Team	**F1 Score**	**Cohen's Kappa**	**Score**
YujinHu	0.845	0.845	0.845
Nitinsinghal	0.792	0.792	0.792
Ternaus	0.789	0.789	0.789
Zhangjingmri	0.778	0.778	0.778
sdsy888	0.759	0.759	0.759
cvblab	0.757	0.757	0.757
XiaHua	0.716	0.716	0.716
AlirezaFatemi	0.712	0.712	0.712
Jpviguerasguillen	0.649	0.64	0.64
qq604395564	0.643	0.643	0.643
Unipabs	0.587	0.587	0.587
Our Proposed	**0.722**	**0.73**	**0.732**

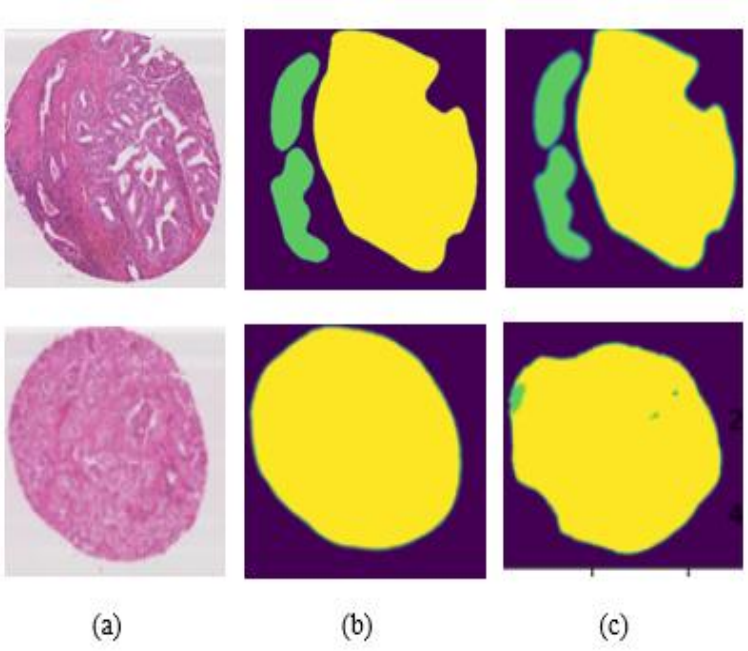

Figure 5.5: Results on best performing model UNET-ResNet50 (a) Original TMA images (b) Original masks (c) Predicted

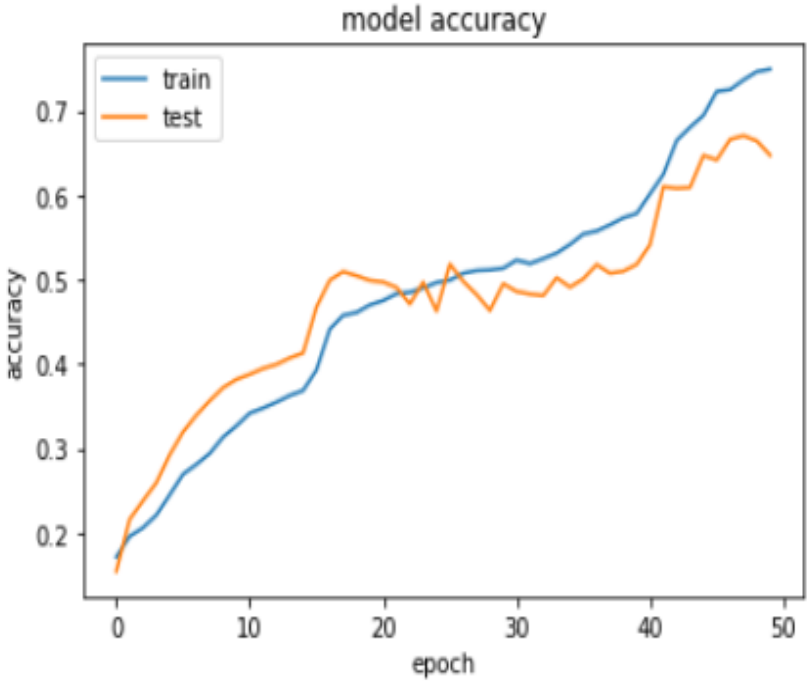

Figure 5.6: Graphical Representation of ResNet50 accuracy on MICAAI dataset

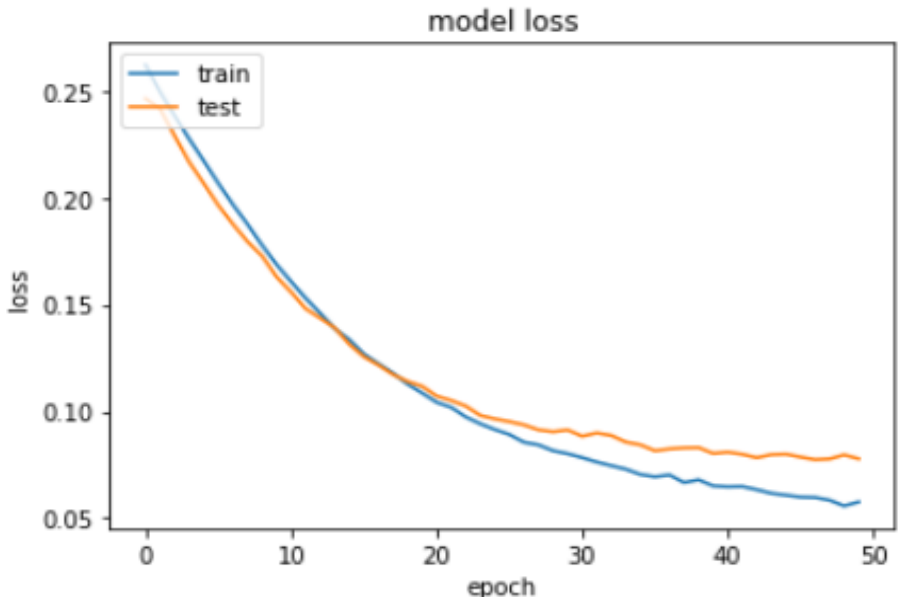

Figure 5.7: Graphical Representation of ResNet50 loss on MICAAI dataset

Figure 5.5 (a) shows original images of MICCAI dataset, (b) shows the original masks which contains pixels ranging from 0 to 7 and (c) shows the predicted masks based on our best performing architecure. **Figure 5.6** and **Figure 5.7** show the training and testing accuracy of ResNet50 model with their corresponding losses. We trained our model on 50 epochs. The graphs show that training as well as testing accuracy gradually increases as number of epochs increases. However, the training and testing loss gradually decreases as the number of epochs increases due to increment of learning.

In Gleason Challenge 2019 [5], teams from all over the world have participated and are working on MICCAI dataset. **Table 5.4** shows the comparison of top selected teams taken from Kaggle in Gleason Challenge with our top performing model UNET-ResNet50. Most of the teams have managed to achieve good results. However, their methodology and results have not been published so far. We have still managed to achieve competitive results on the Gleason Challenge dataset.

5.3 Results on Harvard Dataset

In Harvard dataset, four different CNN architectures are used as our UNET model encoder. We have achieved good results on all four encoder architectures and evaluated those using Dice Score, Cohen's kappa and F score. ResNet50 has performed well as compared to other encoder architectures due to its residual property and faster convergence. Categorical cross entropy is used as loss function with learning rate of 0.0001. ResNet50 has achieved overall score of 0.728 which is highest among other encoders.

5.3.1 Comparison of Results on Harvard Dataset

Table 5.5 shows the results of Harvard dataset. On Harvard dataset, we have implemented VGG19, ResNext50, MobileNetV2 and ResNet50 architectures as encoder with UNET. It is readily apparent from **Table 5.5** that ResNet50 is again the top model in comparison to other encoders with regard to all variables. Dice score of ResNet50 is greater by 2.6%, while Cohen's score by 5.5% from VGG-19, hence it attained the highest accuracy among the four models. ResNet50 stands out from the rests in terms of overall score as well, with 5.5% increase from VGG-19 and Mobile Net and 7% greater than ResNext50.

Table 5.5: COMPARISON OF OUR MODEL WITH BEST PERFORMING MODEL IN TERM OF COHEN'S KAPPA ON HARVARD DATAVERSE DATASET

UNET Model	Encoder Backbone	Dice Score	Cohen's Kappa	Overall Score
1	VGG-19	0.75	0.69	0.69
2	ResNext 50	0.62	0.68	0.68
3	MobileNETV2	0.70	0.63	0.69
4	**ResNet 50**	**0.77**	**0.73**	**0.73**

ResNet50 contains 48 convolution layers stacked one after the other, with max and average pooling. It is a long deep trained model with residual block which gives us state of the art results as compared to other models. Due to identity mapping property, gradient loss problem is solved and network learning speed increased. That is why we have achieved best results with UNET-ResNet50 architecture.

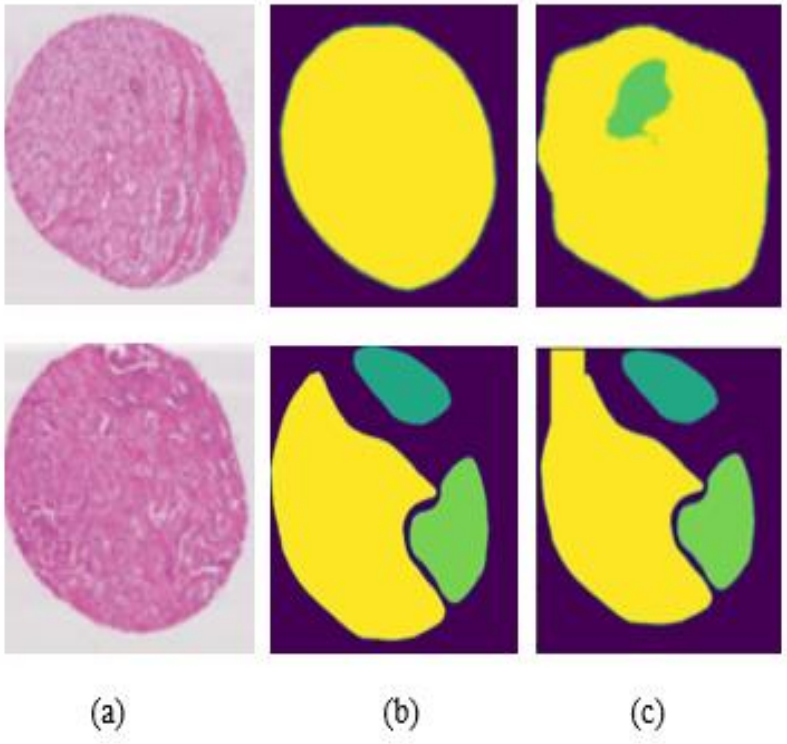

Figure 5.8: Results on best performing model UNET-ResNet50 (a) Original TMA images (b) Original masks (b) Predicted masks

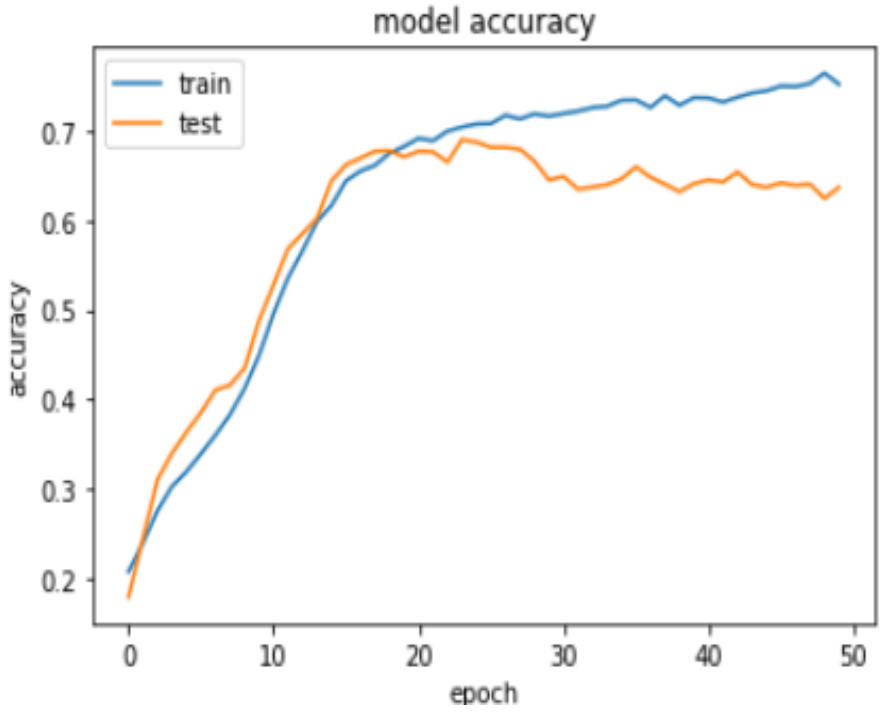

Figure 5.9: Graphical Representation of ResNet50 accuracy on Harvard dataset

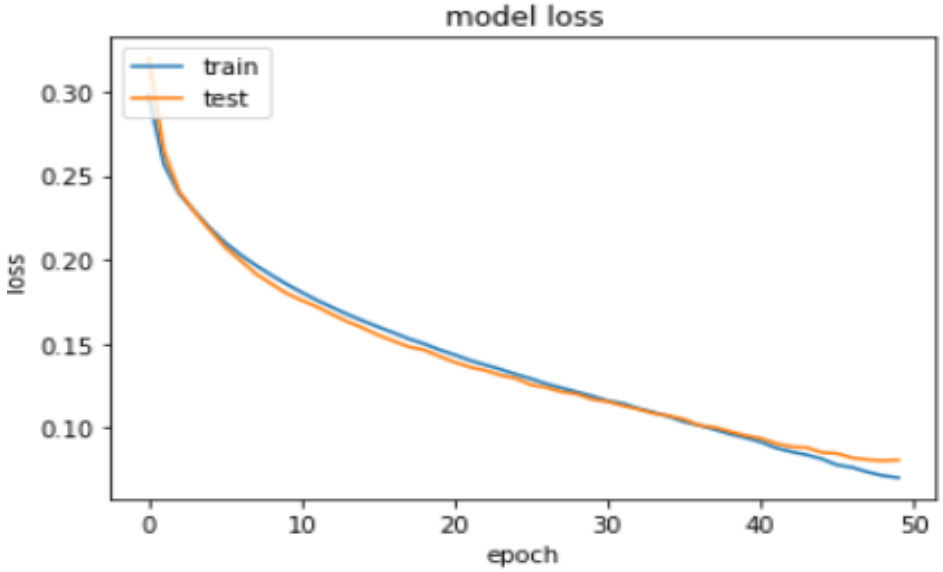

Figure 5.10: Graphical Representation of ResNet50 accuracy on Harvard dataset

Figure 5.8(a) shows the example images from the dataset, (b) shows the ground truth masks of Harvard dataset which contains pixels' ranges from 0 to 4 and (c) shows the predicted masks using

our best performing architecture. **Figure 5.9** and **Figure 5.10** show the training and testing accuracy with corresponding loss. As the number of training epochs are increased, the training accuracy is increased and loss decreases. We tested our model using same parameters and achieved state of the art results. In testing our model generalization, we have used same dataset as used in given [2] and achieved better results as compared to them. The detailed comparison of our results with other results reported in literature is shown in **Table 5.6**.

Table 5.6: COMPARISON OF UNET BASED MODEL RESULTS ON HARVARD DATASET

Model	No. of Images	Cohen's Score
Eirini et al. [2]	640	0.72
Bulten [66]	**640**	**0.71**
UNET-ResNet 50	**640**	**0.73**

5.4 Hardware Requirements

The above experiment was carried out on an Intel Core i5 processor using an NVIDIA GPU with 8 GB of RAM. NVIDIA began producing Computation Unified System Architecture (CUDA) in 2007 to enable programming for the NVIDIA graphical processing unit (GPUs) on Google Colab. Since then, NVIDIA GPUs have become the industry leader with high-performance GPUs specially built for greater matrix multiplication that can be multiplied.

The area of machine learning has achieved a major performance improvement with the introduction of a high-performance graphical processing unit as deep learning techniques involve a massive amount of matrix multiplication and these GPUs have enabled it. Before that, the graphic performance of these GPUs was solely dependent on the capacity of the CPUs. The performance power of the previous NVIDIA GPU version was up to 70 times more than the ordinary personal computers. Since the release of the previous NVIDIA GPU model in 2012, both the memory space and the number of NVIDIA graphics card cores have increased.

6 Chapter 6: Conclusion and Future Work

The results prove that deep learning models have the capability to achieve expert level results. In this thesis, we have implemented deep learning based models on two different datasets for automatic prostate cancer grading. We have proposed a methodology which is based on UNET model for automatic prostate cancer grading at pixel level and predicted the pathologist's level results on both datasets. We used four different CNN architectures; VGG19, ResNext50, MobileNetV2 and ResNet50 as an encoder to UNET model. UNET with ResNet50 encoder gave us state of the art results as compared to other encoder architectures due to its unique identity mapping. Due to lesser number of samples, we have also implemented data augmentation on both datasets which increases the overall performance of UNET model. Our experimental results show that our proposed deep learning based model achieved good results on Gleason Challenge dataset and higher results on Harvard dataset as compared to previous reported results.

6.1 Future Work

Most automatic segmentation methods have promising results in prostate cancer identification and grading, however, further improvement in these algorithms and availability of large amount of data with less class imbalance problem may improve these methods and prove to be helpful in the development of large-scale, clinically acceptable methods for grading of prostate cancer. Moreover, researchers should focus on large scale CNN architectures for more robustness of the model and better generalization.

The majority of automated computer-based methods for prostate cancer grading and analysis have promising results. However, further improvements in these algorithms and the availability of additional image details from new image modalities can enhance these methods and can be useful in the creation of large-scale, clinically appropriate methods for better automatic prostate cancer grading. The Deep Convolutional Neural Network performs better with larger dataset. Therefore, in future the availability of more training samples could be beneficial for network training. Another important factor to be considered when designing an automated prostate cancer grading model is its robustness, which means that the algorithm should be able to deal with a variety of datasets. Hence, in order to design a robust automated prostate cancer grading model,

the availability of more number of datasets would help researchers test their algorithm on various datasets with various modalities.

References

[1] W. Li, J. Li, K. V. Sarma, K. C. Ho, S. Shen, B. S. Knudsen, *et al.*, "Path R-CNN for prostate cancer diagnosis and gleason grading of histological images," *IEEE transactions on medical imaging*, vol. 38, pp. 945-954, 2018.

[2] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, *et al.*, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Scientific reports*, vol. 8, pp. 1-11, 2018.

[3] S. F. Faraj, S. M. Bezerra, K. Yousefi, H. Fedor, S. Glavaris, M. Han, *et al.*, "Clinical validation of the 2005 ISUP Gleason grading system in a cohort of intermediate and high risk men undergoing radical prostatectomy," *PloS one*, vol. 11, 2016.

[4] J. Gordetsky and J. Epstein, "Grading of prostatic adenocarcinoma: current state and prognostic implications," *Diagnostic pathology*, vol. 11, p. 25, 2016.

[5] M. G. C. f. Pathology. (17 October 2019). *Gleason 2019 Challenge*. Available: https://gleason2019.grand-challenge.org/

[6] H. J. Lavery and M. J. Droller, "Do Gleason patterns 3 and 4 prostate cancer represent separate disease states?," *The Journal of urology*, vol. 188, pp. 1667-1675, 2012.

[7] C. C. Huang, M. X. Kong, M. Zhou, A. B. Rosenkrantz, S. S. Taneja, J. Melamed, *et al.*, "Gleason score 3+ 4= 7 prostate cancer with minimal quantity of gleason pattern 4 on needle biopsy is associated with low-risk tumor in radical prostatectomy specimen," *The American journal of surgical pathology*, vol. 38, pp. 1096-1101, 2014.

[8] P. A. Humphrey, "Gleason grading and prognostic factors in carcinoma of the prostate," *Modern pathology*, vol. 17, pp. 292-306, 2004.

[9] M. T. Farooq, A. Shaukat, 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)*, 2017, pp. 642-645.

[10] Q. Zhu, B. Du, and P. Yan, "Boundary-weighted domain adaptive neural network for prostate MR image segmentation," *IEEE transactions on medical imaging*, vol. 39, pp. 753-763, 2019.

[11] Y. Wang, B. Zheng, D. Gao, and J. Wang, "Fully convolutional neural networks for prostate cancer detection using multi-parametric magnetic resonance images: an initial investigation," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 3814-3819.

[12] C. Genestie, B. Zafrani, B. Asselain, A. Fourquet, S. Rozan, P. Validire, *et al.*, "Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: major importance of the mitotic count as a component of both grading systems," *Anticancer research*, vol. 18, pp. 571-576, 1998.

[13] Dana-Farber Cancer Institute, "What is Carcinoma of the Prostate?," 23-4-2019 2019.

[14] M. Matthew Hoffman, "Picture of the Prostate," 2020.

[15] C.-Y. Kuo, S.-S. Wang, C.-H. Lee, and Y.-C. Ou, "Rare presentation of discrete ureteral metastasis from prostate adenocarcinoma," *Formosan Journal of Surgery*, vol. 49, pp. 242-244, 2016.

[16] J. Cuzick, M. A. Thorat, G. Andriole, O. W. Brawley, P. H. Brown, Z. Culig, *et al.*, "Prevention and early detection of prostate cancer," *The lancet oncology*, vol. 15, pp. e484-e492, 2014.

[17] E. Broaddus, A. Topham, and A. D. Singh, "Incidence of retinoblastoma in the USA: 1975–2004," *British Journal of Ophthalmology*, vol. 93, pp. 21-23, 2009.

[18] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal & Image Processing*, vol. 3, p. 23, 2012.

[19] G. P. Murphy and W. F. Whitmore Jr, "A report of the workshops on the current status of the histologic grading of prostate cancer," *Cancer*, vol. 44, pp. 1490-1494, 1979.

[20] E. Short, A. Y. Warren, and M. Varma, "Gleason grading of prostate cancer: a pragmatic approach," *Diagnostic Histopathology*, vol. 25, pp. 371-378, 2019.

[21] D. F. Gleason and G. T. Mellinger, "Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging," *The Journal of urology*, vol. 111, pp. 58-64, 1974.

[22] J. I. Epstein, M. J. Zelefsky, D. D. Sjoberg, J. B. Nelson, L. Egevad, C. Magi-Galluzzi, *et al.*, "A contemporary prostate cancer grading system: a validated alternative to the Gleason score," *European urology*, vol. 69, pp. 428-435, 2016.

[23] N. Chen and Q. Zhou, "The evolving Gleason grading system," *Chinese Journal of Cancer Research*, vol. 28, p. 58, 2016.

[24] P. M. Pierorazio, P. C. Walsh, A. W. Partin, and J. I. Epstein, "Prognostic G leason grade grouping: data based on the modified G leason scoring system," *BJU international,* vol. 111, pp. 753-760, 2013.
[25] J. I. Epstein, W. C. Allsbrook Jr, M. B. Amin, L. L. Egevad, and I. G. Committee, "The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma," *The American journal of surgical pathology,* vol. 29, pp. 1228-1242, 2005.
[26] O. N. Kryvenko and J. I. Epstein, "Changes in prostate cancer grading: Including a new patient-centric grading system," *The Prostate,* vol. 76, pp. 427-433, 2016.
[27] T. A. Ozkan, A. T. Eruyar, O. O. Cebeci, O. Memik, L. Ozcan, and I. Kuskonmaz, "Interobserver variability in Gleason histological grading of prostate cancer," *Scandinavian journal of urology,* vol. 50, pp. 420-424, 2016.
[28] K. Nagpal, D. Foote, Y. Liu, P.-H. C. Chen, E. Wulczyn, F. Tan, *et al.*, "Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer," *NPJ digital medicine,* vol. 2, pp. 1-10, 2019.
[29] G. Campanella, V. W. K. Silva, and T. J. Fuchs, "Terabyte-scale deep multiple instance learning for classification and localization in pathology," *arXiv preprint arXiv:1805.06983,* 2018.
[30] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. Nagtegaal, I. Kovacs, *et al.*, "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Scientific reports,* vol. 6, p. 26286, 2016.
[31] M. Zare, F. Ghodsbin, I. Jahanbin, A. Ariafar, S. Keshavarzi, and T. Izadi, "The effect of health belief model-based education on knowledge and prostate cancer screening behaviors: a randomized controlled trial," *International journal of community based nursing and midwifery,* vol. 4, p. 57, 2016.
[32] N. Bilgili and Y. Kitis, "Prostate Cancer screening and health beliefs: a Turkish study of male adults," *Erciyes Medical Journal,* vol. 41, pp. 164-170, 2019.
[33] K. Koo, C. D. Brackett, E. H. Eisenberg, K. A. Kieffer, and E. S. Hyams, "Impact of numeracy on understanding of prostate cancer risk reduction in PSA screening," *Plos one,* vol. 12, p. e0190357, 2017.
[34] S. Beck, B. Borutta, U. Walter, and M. Dreier, "Systematic evaluation of written health information on PSA based screening in Germany," *PloS one,* vol. 14, p. e0220745, 2019.
[35] M. Obana and H. O'LAWRENCE, "Prostate cancer screening: PSA test awareness among adult males," *Journal of health and human services administration,* pp. 17-43, 2015.

[36] F. H. Schröder, I. van der CRUIJSEN-KOETER, H. J. de KONING, A. N. VIS, R. F. HOEDEMAEKER, and R. KRANSE, "Prostate cancer detection at low prostate specific antigen," *The Journal of urology,* vol. 163, pp. 806-812, 2000.
[37] W. H. Organization, "Global Health Observatory. Geneva: World Health Organization; 2018," ed, 2018.
[38] O. Gersten and J. R. Wilmoth, "The cancer transition in Japan since 1951," *Demographic Research,* vol. 7, pp. 271-306, 2002.
[39] A. R. Omram, "The epidemiologic transition: a theory of the epidemiology of population change," *Bulletin of the World Health Organization,* vol. 79, pp. 161-170, 2001.
[40] M. M. Center, A. Jemal, J. Lortet-Tieulent, E. Ward, J. Ferlay, O. Brawley, *et al.*, "International variation in prostate cancer incidence and mortality rates," *European urology,* vol. 61, pp. 1079-1092, 2012.
[41] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians,* vol. 68, pp. 394-424, 2018.
[42] M. B. Culp, I. Soerjomataram, J. A. Efstathiou, F. Bray, and A. Jemal, "Recent global patterns in prostate cancer incidence and mortality rates," *European urology,* vol. 77, pp. 38-52, 2020.
[43] T. Tarver, "Cancer facts & figures 2012. American cancer society (ACS) Atlanta, GA: American Cancer Society, 2012. 66 p., pdf. Available from," ed: Taylor & Francis, 2012.
[44] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro,* 2008, pp. 284-287.
[45] P.-W. Huang and C.-H. Lee, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE transactions on medical imaging,* vol. 28, pp. 1037-1050, 2009.
[46] G. Nir, S. Hor, D. Karimi, L. Fazli, B. F. Skinnider, P. Tavassoli, *et al.*, "Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts," *Medical image analysis,* vol. 50, pp. 167-180, 2018.
[47] Y. Smith, G. Zajicek, M. Werman, G. Pizov, and Y. Sherman, "Similarity measurement method for the classification of architecturally differentiated images," *Computers and Biomedical Research,* vol. 32, pp. 1-12, 1999.

[48] R. Farjam, H. Soltanian-Zadeh, R. A. Zoroofi, and K. Jafari-Khouzani, "Tree-structured grading of pathological images of prostate," in *Medical Imaging 2005: Image Processing*, 2005, pp. 840-851.
[49] K. Nguyen, B. Sabata, and A. K. Jain, "Prostate cancer grading: Gland segmentation and structural features," *Pattern Recognition Letters*, vol. 33, pp. 951-961, 2012.
[50] J. Ren, E. T. Sadimin, D. Wang, J. I. Epstein, D. J. Foran, and X. Qi, "Computer aided analysis of prostate histopathology images Gleason grading especially for Gleason score 7," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 3013-3016.
[51] N. Zarei, A. Bakhtiari, P. Gallagher, M. Keys, and C. MacAulay, "Automated prostate glandular and nuclei detection using hyperspectral imaging," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 1028-1031.
[52] A. Paul and D. P. Mukherjee, "Gland segmentation from histology images using informative morphological scale space," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 4121-4125.
[53] T. H. Nguyen, S. Sridharan, V. Macias, A. Kajdacsy-Balla, J. Melamed, M. N. Do, *et al.*, "Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning," *Journal of biomedical optics*, vol. 22, p. 036015, 2017.
[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
[55] 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*, 2016, pp. 770-778.
[56] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, 2015, pp. 448-456.
[57] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[58] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, pp. 436-444, 2015.
[59] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234-241.

[60] Q. Zhu, B. Du, B. Turkbey, P. L. Choyke, and P. Yan, "Deeply-supervised CNN for prostate segmentation," in *2017 international joint conference on neural networks (IJCNN)*, 2017, pp. 178-184.
[61] A. Chaddad, M. J. Kucharczyk, C. Desrosiers, I. P. Okuwobi, Y. Katib, M. Zhang, *et al.*, "Deep radiomic analysis to predict gleason score in prostate cancer," *IEEE Access*, vol. 8, pp. 167767-167778, 2020.
[62] D. Karimi, G. Nir, L. Fazli, P. C. Black, L. Goldenberg, and S. E. Salcudean, "Deep Learning-Based Gleason grading of prostate cancer from histopathology Images—Role of multiscale decision aggregation and data augmentation," *IEEE journal of biomedical and health informatics*, vol. 24, pp. 1413-1426, 2019.
[63] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi, and P. W. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human pathology*, vol. 35, pp. 1121-1131, 2004.
[64] A. A. Khani, S. A. F. Jahromi, H. O. Shahreza, H. Behroozi, and M. S. Baghshah, "Towards automatic prostate Gleason grading via deep convolutional neural networks," in *2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2019, pp. 1-6.
[65] A. Lokhande, S. Bonthu, and N. Singhal, "Carcino-Net: A Deep Learning Framework for Automated Gleason Grading of Prostate Biopsies," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1380-1383.
[66] W. Bulten, H. Pinckaers, H. van Boven, R. Vink, T. de Bel, B. van Ginneken, *et al.*, "Automated gleason grading of prostate biopsies using deep learning," *arXiv preprint arXiv:1907.07980*, 2019.
[67] M. Lucas, I. Jansen, C. D. Savci-Heijink, S. L. Meijer, O. J. de Boer, T. G. van Leeuwen, *et al.*, "Deep learning for automatic Gleason pattern classification for grade group determination of prostate biopsies," *Virchows Archiv*, vol. 475, pp. 77-83, 2019.
[68] H.-K. Shin, S.-H. Hong, Y.-J. Choi, Y.-G. Shin, S. Park, and S.-J. Ko, "Self-Attentive Normalization for Automated Gleason Grading System," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*, 2020, pp. 1101-1105.
[69] U. Ali, A. Shaukat, M. Hussain, J. Ali, K. Khan, M. Khan, *et al.*, "Automatic cancerous tissue classification using discrete wavelet transformation and support vector machine," *J. Basic. Appl. Sci. Res*, vol. 6, pp. 15-23, 2016.
[70] J. Peng, S. Kang, Z. Ning, H. Deng, J. Shen, Y. Xu, *et al.*, "Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging," *European radiology*, vol. 30, pp. 413-424, 2020.

[71] "Review: VGGNet — 1st Runner-Up (Image Classification), Winner (Localization) in

ILSVRC 2014.."
[72] S. P. Mohanty, J. Czakon, K. A. Kaczmarek, A. Pyskir, P. Tarasiewicz, S. Kunwar, *et al.*, "Deep Learning for Understanding Satellite Imagery: An Experimental Survey," *Frontiers in Artificial Intelligence*, vol. 3, 2020.
[73] U. Seidaliyeva, D. Akhmetov, L. Ilipbayeva, and E. T. Matson, "Real-Time and accurate drone detection in a video with a static background," *Sensors*, vol. 20, p. 3856, 2020.
[74] M. L. McHugh, "Interrater reliability: the kappa statistic," *Biochemia medica*, vol. 22, pp. 276-282, 2012.
[75] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. Tempany, M. R. Kaus, S. J. Haker, *et al.*, "Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports," *Academic radiology*, vol. 11, pp. 178-189, 2004.