# Magnetic Resonance Images Based Cervical Cancer Classification Using Convolutional Neural Network

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Biomedical Engineering

By:

Kidist Kebede Regassa

Centre of Biomedical Engineering

Addis Ababa Institute of Technology

Addis Ababa University

Advisor: Dawit Assefa Haile (PhD)

June, 2022

Addis Ababa, Ethiopia

## Declaration

I, the undersigned, declare that this MSc thesis is my original work, has not been presented for fulfillment of a degree in this or any other University, and all sources and materials used for the thesis have been acknowledged.

Name: Kidist Kebede Regassa

Signature: _____

Date: _____

This MSc. thesis has been submitted for examination with my approval as the thesis advisor.

_____

Dawit Assefa Haile (PhD)

**Addis Ababa University**

**School of Graduate Studies**

***Certificate of Examination***

This is to certify that the thesis prepared by Kidist Kebede Regassa entitled: *Magnetic Resonance Images based Cervical Cancer Classification Using Convolutional Neural Network* submitted in partial fulfillment of the requirements for the degree of Master of Science in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Examiner _____Signature _____ Date _____
(Internal)

Examiner _____Signature _____ Date _____
(External)

Advisor Dawit Assefa Haile (PhD) Signature _____ Date _____

_____

Associate Director, Post Graduate Programs

# Abstract

Cervical cancer is one type of cancer which affects the cells of the lower parts of the uterus that connects to the vagina called cervix. According to statistics from the ICO/IARC HPV information center, the yearly estimate of cervical cancer in 2020 was 604, 127 new cases and 341, 831 deaths globally, and the second most common in developing countries. It is also reported that more than 85% of cervical cancer patients are living in low resource setting countries being a major cause of morbidity and mortality. To reduce the death rate, effective screening, diagnosis, staging and treatment should be available. The current thesis study aims to bring automatic staging scheme to be used as a decision support system for cervical cancer treatment and prognosis. Pelvic Magnetic Resonance Images were acquired from St. Paul Hospital Medical Millennium College and Tikur Anbessa Specialized Hospital. Basic image pre-processing was employed on the acquired input images. Then two dimensional convolutional neural network was utilized as an integrated feature extraction and classification scheme. The effect of Network layer variations and important network hyper parameters, including learning rate, number of filters, kernel size and epoch was investigated. The performance of the proposed algorithm in binary (two class) and multi (three and five class) classification were tested and resulted in best classification accuracy of 85%, 68.8% and 56.9% respectively. CNN performance was also compared against two other machine learning approaches, namely Support Vector Machine and K Nearest Neighborhood, where both employed region descriptors as well as Gray level co-occurrence matrix during feature extraction. Results showed that the proposed Neural Network based classification scheme outperforms the two machine learning approaches showing its great promises to assist physicians as a decision support system.



**Key Words:** Cervical Cancer, Convolutional Neural Network, Magnetic Resonance Image, Staging, Classification.

# Acknowledgement

First and foremost, I would like to praise the Almighty God. Without his assistance, this task would not have been possible. Then, my sincere gratitude goes to my advisor, Dr. Dawit Assefa Haile, for his valuable suggestions, directions and encouraging guidance. I couldn't leave without expressing my heartfelt gratitude to Dr. Walid Ali and Dr. Webalem Bedilu, from St. Paul Hospital Millennium Medical College, as well as Dr. Semira Abrar from Tikur Anbessa Specialized Hospital, for their supervision during data collection. Many thanks to the entire staff of the Center of Biomedical Engineering at the Addis Ababa Institute of Technology for their encouragement throughout. Last but not least, my deepest gratitude goes to my family and friends for their unconditional love and support. Thank you all for your support and encouragement.

## Table of Contents

## List of Figures



## List of Tables

**Acronyms**

| | |
|---|---|
| AI | Artificial Intelligence |
| ANN | Artificial Neural Network |
| AUC | Area under Curve |
| CC | Cervical Cancer |
| CNN | Convolutional Neural Network |
| CT | Computed Tomography |
| DICOM | Digital Imaging and Communications in Medicine |
| DWI | Diffusion Weighted Image |
| FIGO | Federation of Gynecology and Obstetrics |
| FLAIR | Fluid Attenuated Inversion Recovery |
| GA | Genetic Algorithm |
| GLCM | Grey-level Co-occurrence Matrix |
| HPV | Human Papillomavirus |
| KNN | K Nearest Neighbor |
| LBP | Local Binary Pattern |
| LINAC | Linear Accelerator |
| LMICs | Low and Middle Income Countries |
| MRI | Magnetic Resonance Imaging |
| PACS | Picture Archiving and Communication System |
| PET | Positron Emission Tomography |
| ROC | Receiver Operating Curve |
| SVM | Support Vector Machine |
| TASH | Tikur Anbessa Specialized Hospital |

# Chapter One

# Introduction

## 1.1. Background

The Cervix is the lower part of the uterus which is divided from the upper part known as corpus, by fibro muscular junction and it connects the vagina with the main body of the uterus, acting as a gateway between them. It is cylindrical in shape with average length of 3cm and 2cm in diameter. The shape and size of the cervix varies from women to women, indeed in a woman at diverse stage of her life time [1]. As it is shown in Fig. 1.1 below, the cervix is divided into two sections: The Ectocervix is the first element of a gynecologic examination that may be viewed from inside the vagina. The External Os, an opening at the center of the Ectocervix that allows passage between the uterus and the vagina, opens to allow transit between the uterus and the vagina. And the second portion is the Endocervix, or Endocervical Canal, a tunnel that runs through the cervix from the External Os to the uterus. The transformation zone is the overlapping border between the Endocervix and the Ectocervix. The Internal Os is the opening to the uterus [2].

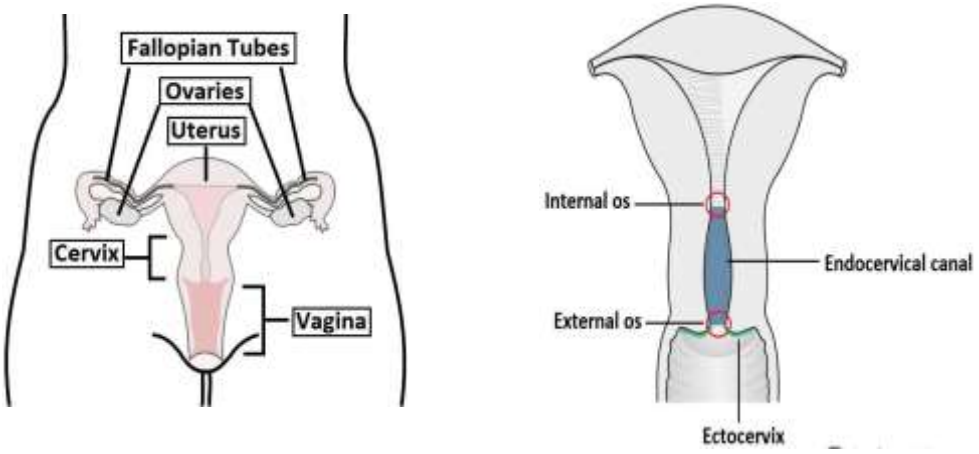



*Figure 1.1: Anatomy of the Cervix [2].*

The cervix serves two primary functions. The first is to allow sperm to enter the uterine cavity. This is accomplished by dilation of the External and Internal Os, and it also serves to maintain the sterility of the upper female reproductive tract. The cervix, as well as all structures above it, are sterile which protect the uterine cavity and upper vaginal tract from bacterial invasion. The frequent shedding of the Endometrium, thick cervical mucus, and a narrow External Os sustain this environment.

### 1.1.1. Cervical Cancer

Cervical cancer (CC) is a form of cancer that affects the cervix's cells. The prevalent histological subtypes are squamous cell carcinoma and adenocarcinoma. The most common variety, squamous cell carcinoma, starts in the thin, flat cells (squamous cells) that line the outer part of the cervix, which spreads into the vaginal canal, whereas adenocarcinoma starts in the column-shaped glandular cells that border the cervical canal. Human papillomavirus (HPV), one of the most frequent types of sexually transmitted virus, is the leading cause of cervical cancer. HPV DNA is found in more than 90% of squamous cervical malignancies. Several studies have shown that HPV 16 and 18 are found in about 70% of all cervical cancers worldwide [3].

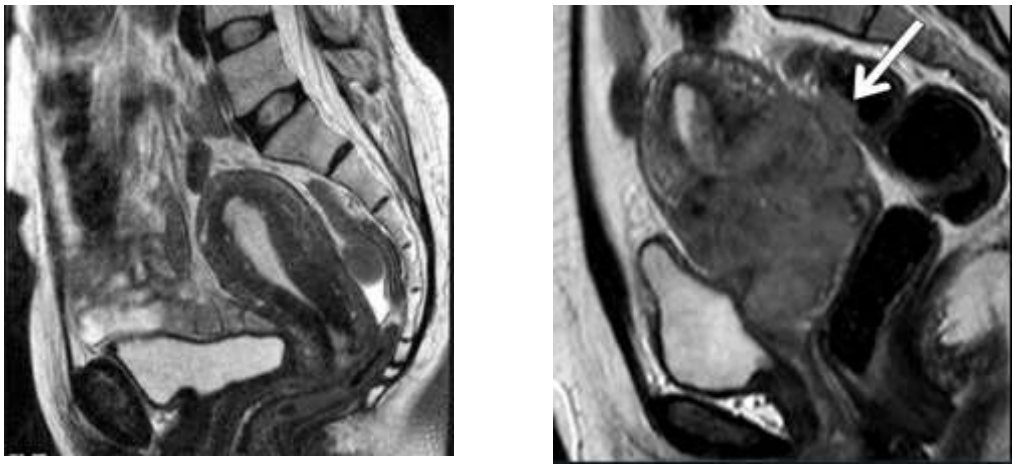


*Figure 1.2: Female Pelvic MRI: (a) Normal, (b) Abnormal [4].*

Figure 1.2 shows a typical sagittal pelvic Magnetic Resonance Image (MRI) for normal and abnormal subjects. The sagittal plane in the midline provides the basis of assessment of most of the parameters required in assessing the positions of the pelvic viscera. This is readily appreciated by depiction of

the fixed structures such as the urethra, cervix, vagina, anorectal junction, levator plate, and the coccyx, which can all be visualized in a single image [5].

### 1.1.2. Current Statistical Data on Cervical Cancer in Africa, Ethiopia

As it is reported, CC affects many women worldwide and more than 85% of CC patients are living in low resource setting countries being a major cause of morbidity and mortality [6]. In Sub-Saharan Africa, 60–75% of women diagnosed with CC reside in rural regions. It is still quite common in Sub-Saharan Africa; rates in underdeveloped nations might be up to 15 times higher than in developed countries [6]. Nonetheless, due to under-reporting, the exact incidence of CC in many African nations is hardly known. Only a few nations have functioning cancer registries, and record-keeping is either non-existent or inadequate. Because most women cannot obtain medical care and die at home, some of the figures published in the literature are hospital-based, which reflects a small fraction of women dying from CC. It is believed CC mortality in Africa is particularly high, owing to the fact that many women who have CC go untreated, due to a lack of access to health care (both financially and geographically).

According to studies, Sub-Saharan Africa accounts for more than 70% of HIV infections. HIV-positive women are more likely to contract HPV at a young age (13–18 years) and at a higher risk for CC [7]. Other risk factors for CC include beginning sexual activity at a young age, having multiple sexual partners (more than one), and polygamous marriage [8]. In some societies, very young virgin girls are married off to much older men, some of whom have three or more wives. This could enhance the chances of a girl contracting HPV during her first intercourse with her husband. Polygamy has been shown to double the risk of CC, with the risk increasing with the number of wives. In general, most women's sociocultural and low socioeconomic level puts them at a higher risk of CC. In addition to the above factors, the country's low economic level, as well as inadequacies in public health policy, has an impact on the spread of CC. One expression of the problem is a lack of awareness and knowledge regarding CC in both rural and urban areas of Sub-Saharan African countries. Most sub-Saharan African countries cannot afford the HPV vaccine due to their low economic status [9].

With almost 112 million people, Ethiopia is Africa's second most populous country behind Nigeria [10]. Since oncology services are proportional to the number of people, the burden of cancer rises as the population grows. According to a report by the International Agency for Research on Cancer Information Center on HPV and Cancer, every year, 7,445 Ethiopian women are diagnosed with CC, with 5,338 dying from the disease [11]. Given the low level of awareness, limited access to screening programs, and the lack of a national cancer registry, these statistics are likely to be lower than the actual number of cases. [12]. It is a severe public health problem that disproportionately affects the country's most vulnerable populations, rural, poor and HIV-positive women, who are more likely to develop advanced cancer stages. As it is shown in Fig. 1.3 below, CC is the second leading cause of cancer deaths among women aged 15 to 44, after breast cancer.

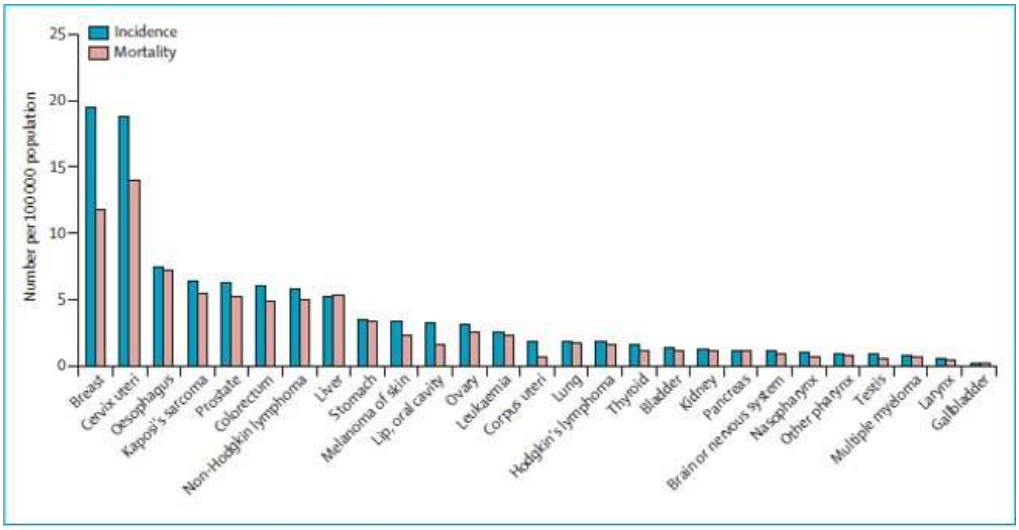



*Figure 1.3: Cancer in Ethiopia [12].*

Most CC patients have a low probability of recovery since they are diagnosed late due to economic, social, and other problems including lack of early detection services, lengthier treatment waiting times, and a lack of understanding about risk factors, signs, and symptoms [13]. CC has enormous financial burden on patients and their families. The cost was determined by the distance between the patient's home and the hospital. Cost was also predicted by the number of employed household members, the number of facilities visited, and the patient's occupation. Higher inpatient costs were linked to longer inpatient hospital stays and the presence of co-morbidity.

The primary goal should be to prevent CC from occurring in the first place, in order to avoid the problem in all aspects, including its financial cost [14].

CC patients also experience significant physical and emotional problems, according to researchers. It has an impact on interpersonal relationships and alters their societal duties and obligations. Furthermore, it has an impact on their sexual lives, causing their relationship with their significant other to deteriorate [15]. Hence, women diagnosed with CC are advised to seek medical, financial, and psychosocial assistance as a remedy. In general, there is a pressing need for improved awareness, primary prevention through vaccination, and early detection programs to minimize CC morbidity and future expenses in terms of both human lives and financial resources. The cancer stage is one of the most important prognostic factors, despite the fact that there are few research addressing predictors of late and advanced stage presentation of CC patients in low- and middle-income countries (LMICs) [16]. Figure 1.4 shows the International Federation of Gynecology and Obstetrics (FIGO) data from the year 2018 that shows number of patients residing in Addis Ababa diagnosed for different stages of CC.

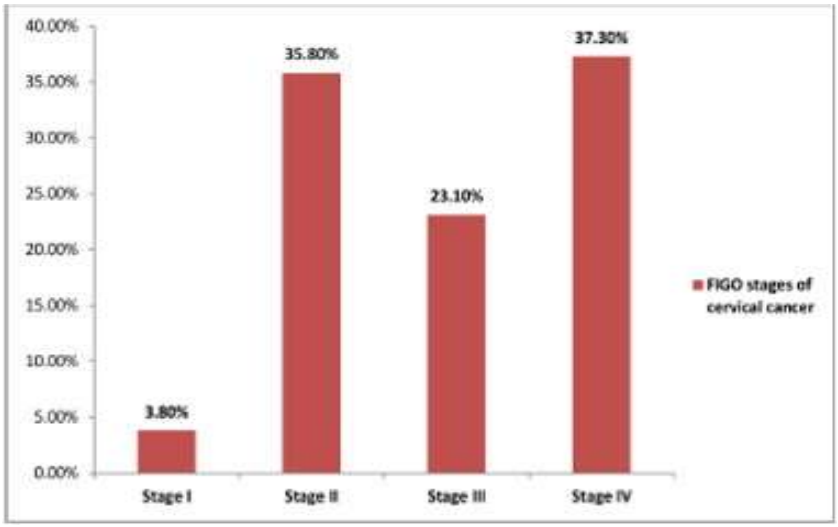



*Figure 1.4: FIGO Stages of Cervical [17].*

Treatment options are dependent on the stage of the cancer. In early stages, a minor surgery can be performed but with advanced cancer stages, radiotherapy (RT) and chemotherapy are common

treatment options. However, in LMICs, the availability of RT services is limited. In Africa, for example, 28 countries do not have access to RT. Even in countries where RT is used, the quantity of RT equipment is grossly insufficient. Furthermore, because Cobalt-60 units are less expensive to install and maintain, they account for 30% of the available RT machines in Africa, rather than intensity-modulated RT using linear accelerators (LINACs), which is the gold standard since it reduces harmful effects from superfluous irradiation of adjacent tissues [18].

Chemotherapy and analgesics, in addition to radiotherapy, must be readily available and affordable. Regional hospitals should be reinforced and capacitated, and price choices for treatment and screening should be subsidized or exempted [19]. The government should formulate policy to address this solution and needs more researches on this critical area.

## 1.2. Statement of the Problem

According to FIGO committee report, "Cervical cancer is the fourth most common cancer in women worldwide and the second most common in Low and Middle Income Countries". It is also reported that more than 85% of CC patients are living in low resource setting countries being a major cause of morbidity and mortality [20]. According to statistics from the ICO/IARC HPV information center, the yearly estimate of cervical cancer in 2020 was 604, 127 new cases and 341, 831 deaths globally. In Ethiopia alone, 7,445 new cases of CC are reported annually with about 5,338 deaths [11]. Given the low level of awareness, limited access to screening programs, and the lack of a national cancer registry, these statistics are likely to be lower than the actual number of cases [12]. In addition, oncology services are inadequate with only one cancer center with radiotherapy unit available at Tikur Anbessa Specialized Hospital (TASH) [21]. There is scarcity of senior-level oncologists, very few are engaged at TASH to serve a population of over 100 million people [22]. The scarcity of professionals increases the workloads of those who are accessible, influencing decision-making and quality of health care services. It also has an impact on patients in different perspectives (economic, social and psychological), since many of them come from rural areas in order to get diagnosed and treated. The traditional way of the staging process involves manual analysis of images of the cervix. These manual assessment of the MRI scans is often subjective, time consuming and hardly repetitive.

Therefore, in order to improve the quality of the service and support professionals in treatment and prognosis decision making process, this thesis aims to bring automatic staging scheme of CCs making use of Convolutional Neural Network (CNN).

## 1.3. Objectives

### 1.3.1. General Objective

The general objective of the study is to classify CCs using CNN based scheme on the acquired Magnetic Resonance Images.

**Specific Objectives**

- To implement preprocessing techniques on the acquired cervical data.
- To understand the effects of the CNN basic hyper parameters.
- To test the performance of the algorithm using known evaluation measures based on useful quantitative matrices.
- To check the efficacy of both binary and multi class classification scheme.
- To compare the algorithm with other commonly used techniques.

## 1.4. Scope of the Research

This research is meant to classify T2 Weighted MR Images to detect and stage CC using a CNN based scheme. Training and testing of the developed system will be carried out based on the data acquired from secondary data sources available at local hospitals in Addis Ababa. However, the clinical validation of the work is beyond the scope of the current study.

## 1.5. Significance of the Study

If the project's output is validated and piloted, it can be used as a decision support system in routine clinical practices. It will considerably assist medical professionals in automating classification of CCs thereby availing a faster diagnosis option for them. The doctors can also use the developed system for a telemedicine application to monitor the decision making process remotely. The system will enhance the quality of healthcare, benefiting both patients as well as the medical professionals.

It will benefit the doctors by minimizing their workload due to higher number of patients in one hospital, which in turn helps to create good working environment and make rational decision. Regarding the patients, it will considerably minimize their costs for transportation and other expenditures. It can also help patients get diagnosed in their local areas through telemedicine application, still preserving their social relationships. Thus, the system could offer significant economic and social benefits.

## 1.6. Organization of the Thesis

The rest of the thesis has been organized into five Chapters. Chapter Two goes into detail on the various cervical cancer screening and diagnosis approaches. The most prevalent diagnosis modality (MRI) is also discussed in depth in this Chapter. In addition, the theorotical background of CNN including its fundamental functions and detail architecture is explained. The third Chapter is a review of literatures on the application of CNN and other machine learning algorithms in diagnosis and classification of CCs. Chapter Four discusses detail on the methodology followed to develop the proposed CC staging system. This chapter combines a method for identifying appropriate parameters with architectural detail design of CNN. It also contains basic concepts of Support Vector Machine and K Nearest Neighborhood algorithms. The outcomes of the suggested scheme are presented and discussed in Chapter Five. Finally, Chpter Six presents the research conclusions, recommendations and topics for further research.

# Chapter Two

# Theoretical Background

Many women die because of late diagnosis and untreated CC. Effective screening, early diagnosis and proper treatment delivery for CC patients can minimize the death rate. Currently available screening tools include Bimanual pelvic examination, Pelvic examination under anesthesia, Biopsy, pap test, HPV typing test and use of different imaging modalities like Colposcopy, X-ray, Lymphangiography, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) - CT scan [23].

## 2.1. Screening and Diagnosis Techniques of Cervical Cancer

The bimanual pelvic examination and sterile speculum examination use a tool called a speculum to maintain the vaginal wall open while identifying abnormal changes in the pelvic organs (cervix, uterus, vagina, ovaries, and other surrounding organs). When a pelvic examination is performed, a Pap test (smear) is also performed. This is done by extracting a sample cell from the cervix's surface. On the sample cells, an HPV type test is performed. HPV16 and HPV18 are more common in cervical cancer patients and can assist confirm a diagnosis. The presence of these two strains of HPV is a strong indication of the existence of cervical cancer. Biopsy is the other method. A tiny quantity of tissue is removed and examined under a microscope. A colposcopy can be used to examine the cervix for abnormal regions. Colposcopy can also be used to guide a cervix biopsy. A specific device known as a colposcope is used during a colposcopy. The colposcope, like a microscope, magnifies the cells of the cervix and vagina. It provides a clinician with an illuminated, enlarged image of the vaginal and cervix tissues. The colposcope isn't put inside the body, and the examination is comparable to that of a speculum. The expert may re-examine the pelvic area while the patient is under anesthesia to check whether the cancer has progressed to any organs surrounding the cervix, such as the uterus, vagina, bladder, or rectum, if it is essential for treatment planning.

X-ray is one of the imaging modalities available. An X-ray uses a little quantity of radiation to generate an image of the structures inside the body. CT scan is a more advanced imaging protocol that also makes use of X-ray. A CT scan uses X-rays captured from various angles to create images of the interior of the body. These images are combined by a computer into a detailed, three-dimensional image that reveal any anomalies or malignancies. A CT scan can be performed to determine the size of a tumor.

The other imaging option is MRI, which produces detailed pictures of the body using magnetic fields rather than X-rays. MRI scan can be performed to determine the tumor size. MRI depicts the distribution of hydrogen nuclei in various tissues and other characteristics linked to hydrogen nuclei mobility in water and lipids by using radiofrequency pulses or signals in the presence of precisely regulated magnetic fields. Improvements in spatial resolution, contrast agents, and most importantly imaging speed have all been made. MRI has proven to be useful in determining the size of the cervix and detecting parametrial invasion, bladder and rectal invasion, the presence and consistency of enlarged lymph nodes, obstruction of the ureter, and lung or liver metastases, thanks to its superior soft tissue contrast resolution and multiplanar capabilities.

A PET scan, is another imaging technique for creating images of organs and tissues inside the body. The patient is given a tiny quantity of a radioactive sugar material to inject into his/her body. The cells that consume the most energy absorb this sugar molecule. Cancer absorbs more of the radioactive material since it uses energy aggressively. The material is then detected by a scanner, which produces pictures of the inside of the body. Lymphangiography has been thoroughly examined in comparison to other modalities for detecting lymph node metastases, all of which are considered developing technologies for diagnosing and evaluating prognoses for CC. Lymphangiography is a procedure that involves directly cannulating the lymphatic ducts and injecting contrast chemicals into the lymph nodes, causing normal lymph nodes to look opaque on films. Ultrasound is another modality which is a real-time, coherent mechanical vibration of sound at very high frequencies. It is produced by piezoelectric materials when a voltage is placed across them, causing a change in thickness.

In gynecologic practice, transabdominal and transvaginal ultrasonography are commonly utilized. Several researchers have utilized Transrectal Ultrasonography (TRUS) to assess tumor size and parametrial involvement. Ultrasonography has a poor contrast resolution, making it difficult to clearly see tumors and distinguish between tumor tissue and surrounding normal tissues. Ultrasonography, on the other hand, is effective in detecting uterine and cervical leiomyomas, which can expand the cervix but are not tumors. Ultrasonography can also be used to assess urinary tract blockage in individuals with advanced illness [24]. CC screening may be made more accurate and accessible with the use of optical technology. Battery-powered digital cameras, for example, may capture multi-spectral pictures of the whole cervix, indicating problematic regions, and high-resolution optical technologies can probe those areas further, allowing for in vivo diagnosis with high sensitivity and specificity. Furthermore, tailored contrast agents can reveal alterations in cervical neoplasia biomarkers. Such advancements could pave the way for a much-needed worldwide CC prevention strategy [24].

In general, among all imaging modalities, MRI can evaluate the actual extent of the disease because of its high spatial and contrast resolution for pelvic tissues and organs. Some advantages of MRI are short acquisition time with multiplanar images, comfort for the patient, absence of ionizing radiation, and mainly, excellent tissue differentiation [25]. Therefore, MRI is employed as an input data to develop the staging system in the current study.

## 2.2. MRI and Its Techniques

MRI is a type of imaging which uses radiofrequency energy to create images of the body. It is advantageous than the others since it doesn't involve any ionizing radiation and it can give multiplanar images by a short acquisition time. Structural details of the body are given in three basic planes: axial (from top to down), sagittal (side to side) and coronal (front to back) [26].

### 2.2.1. Physics of MRI

The magnetic characteristics of atomic nuclei are used in MRI. Figure 2.1 below shows the use of a high, uniform, external magnetic field in order to align the protons that are ordinarily randomly orientated within the water nuclei of the tissue being studied.

This alignment is then interrupted by the introduction of an external Radio Frequency (RF) energy. Through numerous relaxation processes, the nuclei return to their resting alignment and emit RF energy. The emitted signals are measured after a specified interval following the initial RF perturbed by the introduction of an external RF energy. The Fourier transform is used to convert the frequency information in the signal from each position in the scanned plane to matching intensity levels. Different sorts of pictures are produced by changing the sequence of RF pulses delivered and captured.

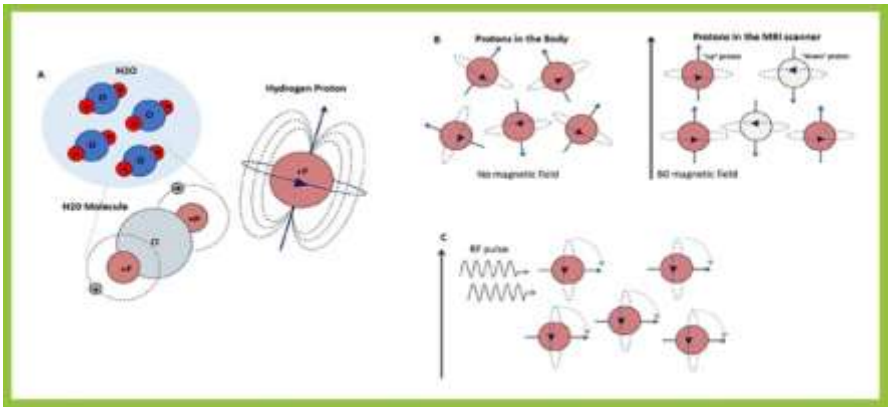


*Figure 2.1: Hydrogen protons and how they behave in a magnetic field [27].*

### 2.2.2. MRI Imaging Sequences

The length of time between subsequent pulse sequences delivered to the same slice is referred to as the Repetition Time (TR). The period between the delivery of the RF pulse and the reception of the echo signal is referred to as the Time to Echo (TE). A tissue can be distinguished by two relaxation times: T1 and T2. T1 (Longitudinal Relaxation) is the temporal constant that governs how quickly excited protons return to equilibrium. It is the time it takes for spinning protons to realign with the external magnetic field. T2 (Transverse Relaxation) is the temporal constant that governs how quickly excited protons approach equilibrium or fall out of phase with one another. T1-weighted and T2-weighted scans are the most frequent MRI sequences. Short TE and TR durations are used to generate T1-weighted images. The image's contrast and brightness are mostly influenced by the T1 characteristics of the tissue. Fat has bright and fluid has dark signal intensities in this type of sequences. T2-weighted images, on the other hand, are created by employing longer TE and TR durations. The contrast and brightness of these pictures are mostly influenced by the T2 characteristics

of the tissue. In contrast to T1 weighted images, Fat has dark and the fluid has bright signal intensities on T2-weighted images. T1- and T2-weighted images may generally be distinguished by examining the Cerebrospinal Fluid (CSF). T1-weighted imaging shows that CSF is black, whereas T2-weighted imaging shows that it is bright [28]. The Fluid Attenuated Inversion Recovery (FLAIR) sequence is a third widely utilized sequence. The Flair sequence is identical to a T2-weighted picture except for the extremely lengthy TE and TR durations. Diffusion weighted imaging (DWI) is another technique used to identify the random movement of water protons. DWI is a very sensitive technique for identifying acute stroke [29]. There are also other parameters that could be generated on an MRI scanner and there are many literatures available discussing such details of MRI that any one might want to consult.

## 2.3. MRI for Cervical Cancer Diagnosis

MR imaging can provide highly accurate information on the exact extent of tumors because of its fine contrast resolution. CCs appear as hyperintense masses on T2-weighted images regardless of histopathologic type [30]. T2-weighted scans, in addition to detecting the main tumor and its size, give great details of the cervical structure and normal uterus. On this sequence, the normal cervical stroma appears as a low-density signal, while around 95% of uterine cervix cancers show as somewhat hyperintense masses in comparison to the surrounding stroma. Prognostic parameters influencing therapy selection and which can be evaluated by gynecological examination can be conducted by MRI with superior accuracy than clinical examinations and CT as staging tools, notably in the parametrial evaluation. Comparative investigations of the three methods (MRI, CT, and clinical examination) revealed that MRI had 92 % accuracy, 78 % for clinical examination, and 70 % for CT [31].

After diagnosis of CC, staging is done. Staging is a procedure conducted in order to know the location and area of the cancer which are very important parameters to decide what kind of treatment is favorable for that particular stage of the disease and to forecast a patient’s chance of recovery and other patient outcomes. There are three commonly known CC staging organizations: FIGO, the American Joint Committee on Cancer (AJCC), and the International Union against Cancer (UICC). FIGO developed one system of staging, and the UICC and AJCC together developed another system, known as the TNM. In the TNM system, cancer is staged according to the extent of the primary tumor (T), lymph node metastasis (N), and distant metastasis (M).

The FIGO classification is based on tumor size, vaginal or parametrial involvement, bladder/rectum extension, and distant metastases [32]. In Ethiopia, the staging system developed by the FIGO is used predominantly [See Appendix B]. This is the latest staging system revised in 2018 which is designed based on the results of physical examinations, imaging scans, and biopsies. It was approved after FIGO Regional Meeting held in Dubai in April 2018 [20]. Based on studies comparing the TNM and FIGO clinical staging systems in terms of usefulness, it was discovered that the FIGO classification system outperforms the TNM system in uterine cervix cancer in terms of simplicity, recording methods, and predictive value [32][33].

## 2.4. Need for Cancer Staging

### A. Treatment

As it is mentioned previously, staging of CC can help the doctor in preparing the treatment plan for the patient. Favorable treatment should be planned depending on the stage of the cancer. If it is at an early-stage, surgery is required if not chemotherapy and/or radiotherapy are needed [34].

### B. Prognosis

The other advantage of staging is to estimate the chance of recovery. Recovery rate is dependent on time of diagnose. Researchers stated that "when cervical cancer is diagnosed at an early stage, the 5-year survival rate for women with invasive cervical cancer is 92%. If CC has invaded surrounding tissues or regional lymph nodes, the 5-year survival rate is 56%. If the cancer has metastasized to other parts of the body such as in liver and bone, the 5-year survival rate is 17%" [34].

## 2.5. Automatic Cervical Cancer Detection

The traditional way of the diagnosis process involves manual analysis of images of the cervix based on different modalities. While MRI based method is favored because it can accurately assess prognostic indicators, e.g., tumor size, parametrial invasion, pelvic sidewall, and lymph node invasion [35]. Manual assessment of the MRI scans is often subjective, time consuming and hardly repetitive. This is where automation techniques come into play. In this regard there are a number of techniques suggested in the literature to help the automation procedure.

Currently favored techniques to help automate the staging process make use of Artificial Intelligence (AI). The most commonly used techniques for classification under the categories of AI are Computer Vision (CV), Genetic Algorithm (GA) and machine learning algorithms including K – Nearest Neighborhood (KNN), Support Vector Machines (SVM), Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) [36]. These techniques were used to design computer based applications which can be used as a clinical decision support system at the point of health care procedures (diagnosis, staging and treatment).

### 2.5.1. Convolutional Neural Network

Advanced computational techniques have been developed recently and played a major role in different sectors. AI is most dominant by replacing the work of humans in different sectors as communication (speech recognition), automotive (self-driving cars) and medicine industry (robotics and automation in clinical care) [37]. AI is a wide phrase that includes machine learning and deep learning. Because deep learning and machine learning are sometimes used interchangeably, it is important to distinguish between the two. Machine learning, deep learning, and neural networks are all subfields of artificial intelligence. However, deep learning is a subfield of machine learning, and neural networks is a subfield of deep learning, as also depicted in Fig. 2.2.

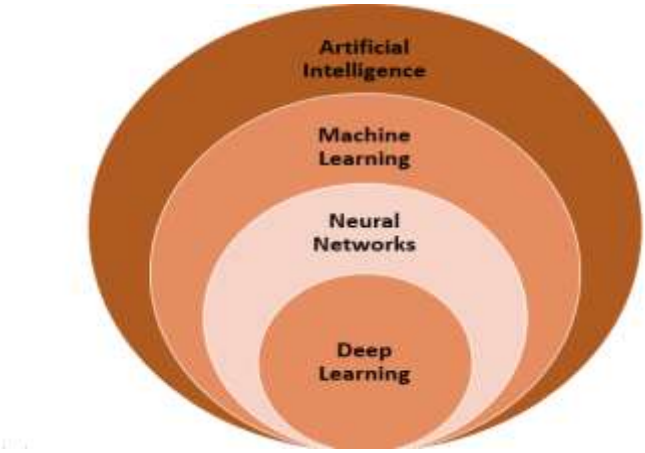



*Figure 2.2: Sub fields of Artificial Intelligence [38].*

Machine learning is the application of data and algorithms to simulate human learning. Machine learning is a crucial part of the rapidly expanding discipline of data science. Algorithms are trained to generate classifications or predictions using statistical approaches, revealing crucial insights in data mining initiatives. Unsupervised learning and supervised learning are the two most extensively used machine learning approaches. The majority of machine learning (about 70%) involves supervised learning. Unsupervised learning accounts for ten to twenty percent of all learnings. Other technologies that are occasionally utilized include semi-supervised and reinforcement learning [39]. Short explanations in each learning types are presented below.

- Supervised machine learning: The use of labeled datasets to train algorithms that reliably classify data or predict outcomes is characterized as supervised learning, often known as supervised machine learning. As more data is introduced into the model, the weights are adjusted until the model is properly fitted. Some common methods of supervised learning are CNN, naive bayes, linear regression, logistic regression, random forest, support vector machine (SVM), and other approaches.
- Unsupervised machine learning: Unsupervised learning analyzes and clusters unlabeled datasets using machine learning methods. Without the need for human intervention, these algorithms uncover hidden patterns or data groupings. Its capacity to find similarities and contrasts in data makes it an excellent choice for exploratory data analysis, cross-selling techniques, consumer segmentation, picture and pattern recognition, and more. A good example of unsupervised machine learning is clustering.
- Semi-supervised learning: Between supervised and unsupervised learning, semi-supervised learning is a good compromise. It guides categorization and feature extraction from a larger, unlabeled data set using a smaller labeled data set during training. Semi-supervised learning can overcome the problem of not having enough labeled data to train a supervised learning algorithm (or not being able to afford to label enough data).
- Reinforcement machine learning: Reinforcement learning is a machine learning training strategy that rewards desirable behaviors while penalizing undesirable ones. A reinforcement learning agent can perceive and comprehend its surroundings, act, and learn through trial and error in general. This method assigns positive values to the desired actions to encourage the agent and

negative values to undesired behaviors. This programs the agent to seek long-term and maximum overall reward to achieve an optimal solution.

The way in which deep learning and machine learning differ is in how each algorithm learns. Deep learning automates much of the feature extraction piece of the process, eliminating some of the manual human intervention required and enabling the use of larger data sets. The “deep” in deep learning is just referring to the depth of layers in a neural network. A neural network that consists of more than three layers, which would be inclusive of the inputs and the output, can be considered a deep learning algorithm or a deep neural network. Deep learning is also known as hierarchical learning since the output from the preceding layer is used as input for each subsequent layer. The network learns a hierarchy of concepts by learning several levels of representations that correspond to different levels of abstraction [37].

ANNs are used in some of the most successful deep learning approaches. The 1959, the biological model proposed by Nobel Laureates David H. Hubel and Torsten Wiesel, who discovered two types of cells in the primary visual cortex (simple cells and complex cells), was the inspiration for ANNs. Many ANNs can be thought of as cascading models of cell types inspired by these biological findings [37]. ANNs are computing systems that are substantially influenced by organic nerve systems (such as the human brain). ANNs are primarily made up of a large number of interconnected computational nodes (also known as neurons) that work together in a dispersed manner to learn from the input and optimize the final output. It is made up of a number of node layers, each of which has an input layer, one or more hidden layers, and an output layer [39]. The input is normally entered into the input layer as a multidimensional vector, which is then distributed to the hidden layers. The hidden layers will then make judgments based on the previous layer’s decisions, weighing how a stochastic change within itself affects or improves the final output, which is known as the learning process. Deep learning is defined as having numerous hidden layers piled on top of each other. Each node, or artificial neuron, is connected to the others and has a weight and threshold linked with it. If a node’s output exceeds a certain threshold value, the node is activated, and data is sent to the next tier of the network [39].

CNNs are similar to classic ANNs in that they will still express a single perceptual score function from the input raw image vectors to the end output of the class score (the weight). The last layer will contain loss functions related with the classes, and all of the standard tips and tricks created for classic ANNs will remain applicable. As it is shown in Fig. 2.3, the only distinction between CNNs and standard ANNs is that CNNs combine automatic feature extraction and discriminative classifiers into a single model, which helps to learn hierarchical representations.

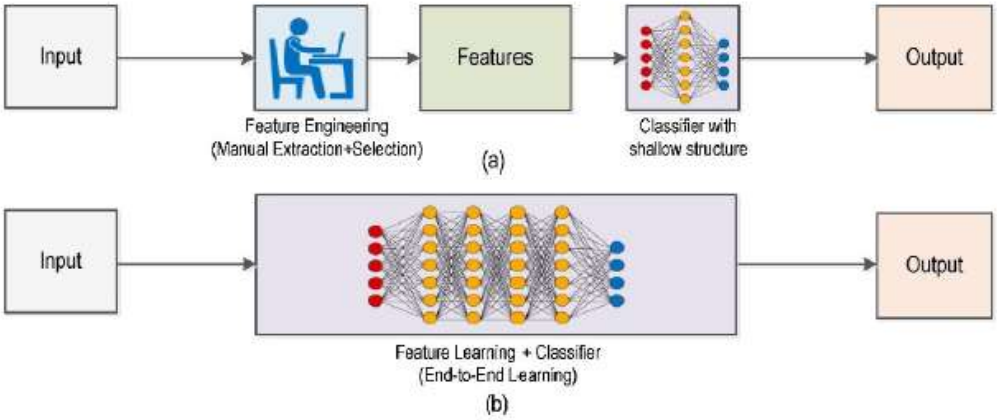



*Figure 2.3: Traditional Method vs. Deep Learning (CNN) workflow [40].*

The conventional CNN is composed of fully linked layers as well as a number of blocks that include convolutions, an activation function layer, and pooling layers [41]. CNN networks outperform classic fully connected neural networks in terms of speed and robustness [42]. CNNs are variants of multilayer perceptrons meant to utilize minimum preprocessing and were inspired by biological processes. They have a wide range of applications, including image and video recognition, classification system, recommender systems, and natural language processing [37]. CNN has also shown to be highly effective in handling image classification issues. CNN-based research greatly increased the best performance for several image databases, including the MNIST, NORB, and CIFAR10 datasets. It excels in inferring local and global structures from image data [42].

**Basic Functionality of CNN**

a. **Input Layer:** will hold the pixel values of the image which is the image data. Image data must be represented by a three-dimensional matrix. It needs to be reshaped in a single column [39].

b. **Convolutional Layer:** As the name refers, the convolutional layer is fundamental in the functionality of CNNs. The layers parameters focus around the use of learnable kernels which are usually small in spatial dimensionality, but moves along the whole depth of the input. When the data and the convolutional layer coincide, the layer convolves each filter across the spatial dimensionality of the input to produce a 2D activation map [39]. In image classification tasks, one or more 2D matrices (or channels) are treated as the input to the convolutional layer and multiple 2D matrices are generated as the output. The number of input and output matrices may be different. The process to compute a single output matrix is defined as [42]:

$$A_j = f(\sum_{i=1}^{N} I_i * K_{i,j} + B_j) \qquad (2.1)$$

where $I_i$ is input matrix, $K_{i,j}$ is the kernel, $B_j$ is a bias, f is a nonlinear activation function and $A_j$ is the output matrix. Firstly each input matrix $I_i$ is convoluted with a corresponding kernel matrix $K_{i,j}$. Then the sum of all convoluted matrices is computed and a bias value $B_j$ is added to each element of the resulting matrix. Finally, a non-linear activation function f is applied to each element of the previous matrix to produce one output matrix $A_j$. Each set of kernel matrices represents a local feature extractor that extracts regional features from the input matrices [42].

c. **Neuron Activation:** A broader classification is made in the form of linear and non-linear activation functions. A linear function passes the input at a neuron to the output without any change. Since, deep network architectures are designed to perform complex mathematical tasks, non-linear activation functions have found wide spread success [43]. In ANN, non-linear transfer functions are used as the neuron activation function. For example, commonly used activation functions include sigmoid function $f(x) = 1/(1 + e^{-x})$ and hyperbolic tangent function $f(x) = \tanh(x)$. Both sigmoid and hyperbolic tangent are saturating non-linear functions that the output gradient drops close to zero as the input increases. Some recent studies suggested that non-saturating non-linear function like rectified linear function or ReLU given by $f(x) = max(0, x)$. ReLU improves both learning speed and classification performance in CNN applications. In most CNN model, the

ReLU activation function is used in the convolutional layer and results showed that the ReLU activation function improves classification performance and the network converges much faster than using the sigmoid activation function [42].

d. **Padding:** Padding, typically zero padding, is a technique to address the issue, where rows and columns of zeros are added on each side of the input tensor, so as to fit the center of a kernel on the outermost element and keep the same in-plane dimension through the convolution operation. The distance between two successive kernel positions is called a stride, which also defines the convolution operation. The common choice of a stride is 1; however, a stride larger than 1 is sometimes used in order to achieve down sampling of the feature maps. An alternative technique to perform down sampling is a pooling operation, which is described below [44].

e. **Pooling Layer:** This will then simply perform down sampling along the spatial dimensionality of the given input, further reducing the number of parameters within that activation. The pooling layer operates over each activation map in the input, and scales its dimensionality using the given pooling function. Pooling layer plays an important role in CNN for feature dimension reduction. In order to reduce the number of output neurons in the convolutional layer, pooling algorithms should be applied to combine the neighboring elements in the convolution output matrices. Commonly used pooling algorithms include max-pooling and average-pooling [42].

f. **Batch Normalization:** A major issue in using deep convolutional network (DCNN) is over-fitting of the model during training. Over fitting refers to a situation where a model learns statistical regularities specific to the training set, i.e., ends up memorizing the irrelevant noise instead of learning the signal, and, therefore, performs less well on a subsequent new dataset. There have been several methods proposed to minimize over fitting. The best solution for reducing over fitting is to obtain more training data. A model trained on a larger dataset typically generalizes better, though that is not always attainable in medical imaging. The other solutions include regularization with dropout or weight decay, batch normalization, and data augmentation, as well as reducing architectural complexity [44]. It has been shown that dropout is used successfully to avoid over-fitting. A dropout layer drops certain unit connections which are selected randomly. Dropout layer is widely used for regularization. In addition to dropout, batch normalization has also been successfully used for the purpose of regularization. The input data is divided into mini batches. It

is shown that using batch normalization not only speeds up the training but, in some cases, preform regularization eliminating the need for using dropout layers [43].

g. **Fully-connected Layers:** This will then perform the same functions as typical ANNs, attempting to generate class scores from the activations for classification. It is also proposed that ReLU be utilized between these layers to boost performance [39].

h. **Last Layer Activation Function:** Typically, the activation function applied to the last fully connected layer differs from the others. Each activity necessitates the selection of an appropriate activation function. A softmax function is used as an activation function in the multiclass classification task to normalize output real values from the last fully connected layer to target class probabilities, where each value ranges between 0 and 1 and all values total to 1 [44].

i. **Training a Network:** Training a network is a process of finding kernels in convolution layers and weights in fully connected layers which minimize differences between output predictions and given ground truth labels on a training dataset. Back propagation algorithm is the method commonly used for training neural networks where loss function and gradient descent optimization algorithm play essential roles. A model performance under particular kernels and weights is calculated by a loss function through forward propagation on a training dataset, and learnable parameters, namely kernels and weights, are updated according to the loss value through an optimization algorithm called back propagation and gradient descent, among others [44].

- *Loss function:* A loss function, also referred to as a cost function, measures the compatibility between output predictions of the network through forward propagation and given ground truth labels. Commonly used loss function for multiclass classification is cross entropy, whereas mean squared error is typically applied to regression to continuous values. A type of loss function is one of the hyper parameters and needs to be determined according to the given tasks [44].
- *Gradient descent:* Gradient descent is commonly used as an optimization algorithm that iteratively updates the learnable parameters, i.e., kernels and weights, of the network so as to minimize the loss. Gradient descent has three types namely Batch Gradient Decent, Stochastic Gradient Decent and Mini – Batch Gradient Decent, each differ in how much data is used to compute the gradient of the objective function. For big datasets, batch gradient descent performs redundant calculations because it recomputes gradients for comparable cases before each parameter change. Stochastic

Gradient Decent eliminates duplication by completing one update at a time. As a result, it is significantly faster. Finally, mini-batch gradient descent takes the best of both kinds and updates it for each mini-batch of n training samples [45].The gradient of the loss function provides us the direction in which the function has the steepest rate of increase, and each learnable parameter is updated in the negative direction of the gradient with an arbitrary step size determined based on a hyper parameter called learning rate.

- *Learning Rate:* The learning rate is a hyperparameter that specifies how much the model should change in response to the predicted error each time the model weights are updated. The learning rate, in particular, is an adjustable hyperparameter used in neural network training that has a tiny positive value, often in the range of 0.0 to 1.0.

CNNs may alter the original input layer by layer utilizing convolutional and down sampling approaches to obtain class scores for classification and regression using this simple method of transformation [39]. One typical issue in real-world implementations of deep learning-based classifiers is that some classes have much more examples in the training set than others. This disparity is known as class imbalance. It influences both model convergence during the training phase and model generalization on the test set. The use of sampling methods is the most easy and popular strategy for solving this problem. These strategies work on the data (rather than the model) to improve balance [41].

As mentioned above, CNNs have recently dominated other traditional algorithms and there are different literatures that showed the application of CNN in the diagnosis and staging of CC. Among those literatures, some of them are reviewed and discussed in the next Chapter.

# Chapter Three

# Literature Review

The literature review part is composed of two sections: the first one discusses the applications of different AI algorithms, other than CNN, in the area of CC and the second focuses on those methods that made use of CNNs in CC studies. Both Sections review the two and multiclass classification of cervical cancer.

## 3.1. Non-CNN based Cervical Cancer Classification

### 3.1.1. Binary Classification

A study by Abid Sarwar et al. discussed performance evaluation of machine learning techniques for screening of cervical cancers [46]. The comparative study was done on fifteen different machine learning algorithms which were implemented on different platforms over two databases in order to evaluate their screening performances for prognosis of cervical cancer based on analysis of pap smear images. Their result showed that ensemble of dichotomies (END) outperformed the rest with prognostic efficiency of about 77 % on the first dataset and about 72% on the second dataset. The worst performer was Naïve Bayes with predictive potential of about 50% on the first dataset and about 60% on the second dataset. Clearly there is ample room to enhance the performance of even the best performing method. Another study by Wasswa William et al. reviewed image analysis and machine learning techniques that are used in the literature for automated cervical cancer screening based on pap-smear images [47]. The authors considered in their review those publications reported in 15 years till the year 2019. The different methods were compared based on images acquired from open pap-smear database. CHAMP digital image software was used as a gold standard. After assessing all the methods and comparing them against the gold stand, the authors finally concluded that KNN algorithm performed best but recommended that combining KNN algorithm with others such as SVM as well as pixel level classifications and incorporating statistical shape models can further improve the performance of the classification. In another research reported previously [48], an automated CC diagnosis system based on textural features and a multiclass SVM based classifier in MRI images was suggested.

The MRI images were first pre-processed to remove noise and other unwanted artifacts. Following pre-processing, the images were segmented using the region growing approach to determine the region of interest. The Grey Level Co-occurrence Matrix (GLCM) was used to extract texture information from the segmented regions. Almost 22 features were retrieved from a segmented area and then fed in to the multiclass SVM classifier to determine whether or not the supplied image contains any malignancy. The proposed technique achieved 94% accuracy in classifying cancerous and non-cancerous cases.

Another research proposed an automated diagnostic system based on image processing techniques, which could help radiologists and play its role in cancer detection [49]. The study described the segmentation and classification algorithms that are used to operate the envisioned automated decision support system based on MR images. Prior to classification, the original strategy was segmentation based on intensity thresholding. Their next stage was feature extraction, which entails minimizing the number of resources required to describe a large quantity of data. The features considered in the study were Pixel List, Major Axis Length, Minor Axis Length, Eccentricity, Orientation and Area. GLCM features such as energy, homogeneity, correlation, and entropy have also been employed. The SVM based classification approach was also compared with a feed forward back propagation based ANN scheme. The study finally reported classification accuracy 92% when using the SVM based scheme while the ANN based scheme offered an accuracy of 84%.

### 3.1.2 Multi Class Classification

Pabitra Mitra et al. published a paper entitled “Staging of Cervical Cancer with Soft Computing’’ [50]. They used hybrid decision support system for detecting the different stages of CC. Their hybrid model includes knowledge-based sub-network modules which utilize the knowledge of the modular structure used for faster convergence, GA using rough set theory to tune the network weights and Interactive Dichotomizer 3 (ID3) algorithm. This integrated model offered overall accuracy of 81.5% during training and 80.2% during testing.

In another research by M. K. Soumya et al., a method that made use of nonlinear SVM classification based on both second-order statistical texture features and transform (Contourlet and Gabor transforms) based features was developed [51]. The GLCM has been widely employed for various texture analysis applications, and the results have been satisfactory [52], [53]. The GLCM functions characterize an image's texture by calculating the frequency with which pairs of pixels with given values and in a specified spatial relationship occur in an image, generating a GLCM, and then extracting statistical measures from this matrix. SVM is used for classification purposes [54]. The authors used MR images as an input to the classifier and the designed classification system resulted in an overall accuracy of 81% when applied on T1-weighted axial images, 82% when applied on T2-weighted axial images and 83% when tested on T2-weighted sagittal images. The classification can predict the patient's stage, whether it is grade I, II, III, or IV.

## 3.2. CNN Based Cervical Cancer Classification

This deep neural network is a new machine learning technology that has shown promises in a variety of classification problems. Notably, CNN outperforms all other image classification algorithms. CNNs are increasingly used to improve clinical practices as diagnosis and treatment procedures. Several studies have demonstrated the use of CNN in the medical field. Yadav and Jadhav [55] illustrated the applicability in a dataset of chest X-rays to classify pneumonia. Gulshan et al., Esteva et al., and Bejnordi et al. proved the potentials of deep learning for diabetic retinopathy screening, skin lesion categorization, and lymph node metastasis identification, respectively. A study reported in [56] carried out experiments to obtain an accurate contour of the bones. The method allowed CNN to provide an exact delimitation of a specific bone and is an entirely different way of employing the network for X-ray pictures. The results were excellent, allowing this form of neural network to be used to extract any contour of the bones from an X-ray image. Schwendicke et al. [57] manifested the application of CNNs for dental image diagnostics. Other key applications of CNN in cancer detection include lung cancer diagnosis [58], breast cancer [59] and brain tumor identification [60].

### 3.2.1 Binary Classification

In a study by Xiao Qing Zhang & Shu-Guang Zhao, a deep-learning model to classify cervical images based on a scheme that involves two steps was developed [61]. First U-Net CNN algorithm was used to segment the cervical images and then the Caps Net cervix network model was applied to classify the segmented images. The method was computationally fast during the training session and offered a classification accuracy of around 80%. The study faced a problem of over fitting which can be reduced by increasing the training data. The study recommended that further adjustment on the design of the Caps Net cervix network model is very crucial in order to get better performance.

Huyen Nguyen et al. proposed another model designed using CNN for classification of cervix images [62]. They built two new models and also adopted two existing pre trained models. After doing comparison on the performance of the models, CervixNet - 2 model outperformed the rest with an overall classification accuracy of 63%, which was way below satisfactory. The authors recommended use of different image preprocessing techniques and advanced image segmentation algorithms for increasing the accuracy of the classification process. The study in [63] provided another fully-automated deep learning pipeline for detecting the cervix region and classifying cervical cancers. The proposed deep learning pipeline is taught and assessed utilizing cervigram pictures from the Guanacaste project led by the American National Cancer Institute (NCI) and the Intel & Mobile ODT Kaggle challenge. The algorithm's speed, precision, and lightweight architecture made it ideal for implementation as a smart device application. It was hoped that in the future, it will improve the perceptual quality of cervigram pictures by decreasing the effects of specular reflection and offer more accurate manual labeling of the cervical RoIs.

Another Study on Classification of cervical images using ResNet101 was published in 2019 [64]. Pre trained CNN was used to solve the two class problem (normal and abnormal). The study was conducted on online database provided by the cancer Imaging Archive (TCIA). They initially split the dataset into training and testing dataset. All the dataset were resized to 224 x 224. The overall classification accuracy was obtained to be 97.27%.

### 3.2.2 Multi Class Classification

A study by Ahmed Ghoneim et al. proposed a cervical cancer classification scheme using CNNs and extreme learning machines [65]. The authors proposed use of CNNs for cervical cancer cell detection and classification. They used cell images as input data and CNNs for feature extraction. Finally, for the classification step, an extreme learning machine (ELM)-based method that is integrated with the CNN was used. Transfer learning and fine-tuning techniques were also applied to enhance the overall classification accuracy. The authors achieved 99.5% accuracy during detection (2-classes: normal vs abnormal) and 91.2% during Seven-class classification (3 normal: Superficial squamous epithelial, Intermediate squamous epithelial and Columnar epithelial and 4 abnormal: Mild squamous non-keratinizing dysplasia , Moderate squamous non- keratinizing dysplasia, Sever squamous non-keratinizing dysplasia and Squamous cell carcinoma in situ intermediate ). Comparing with other research reports, these accuracies were better. As recommendation for future work, the authors suggested to test the same algorithm using other databases and also investigate the performance of other deep learning architectures.

The results of a comprehensive study on deep learning-based Computer Aided Diagnostic methods for categorization of cervical dysplasia using pap smear pictures were presented in another study [66]. All investigations were carried out on a genuine indigenous image database of 1611 images generated at two diagnostic centers. The emphasis was on developing an effective feature vector capable of performing many levels of representation of the features buried in a pap smear image. Deep CNN was utilized for this purpose, followed by feature selection using an unsupervised technique and the Maximal Information Compression Index as a similarity metric. The suggested system was also tested on a publicly available database and compared to two current conventional systems. The experimental results and comparison reveal that the suggested approach performs well in the classification of the pap smears.

An auto-assisted cervical cancer detection system based on a CNN trained on the Cervical Cells database is proposed in another research [67]. The network was trained via transfer learning, with initializing weights derived from training on the ImageNet dataset. Three distinct classifiers were proposed for final classification/screening of cell samples: Softmax regression (SR), SVM, and Gentle

Boost ensemble of decision trees (GEDT). For the 2-class problem, the classification accuracies of SR, SVM, and GEDT were 98.8%, 99.5%, and 99.6%, respectively, while for the 7-class problem, they achieved 97.21%, 98.12%, and 98.85%, respectively. These results revealed that the suggested system outperforms its predecessors under a variety of testing situations. The major drawback with this method was its computational complexity where it took about 4 hours and 30 minutes for two class classification and about 8 hours and 20 minutes for 7 class classification.

In summary, literatures on different algorithms for classification of cervical cancers were reviewed in and these are the basic points that we learned:

- Incorporating robust statistical feature extraction techniques, KNN and SVM are recommended classification algorithms on cervical images.
- T2 weighted images are highly favored for CC diagnosis and classification.
- As MRIs are superior to show more details, studies that employ MRIs are particularly beneficial without the need for any biopsy procedures.
- In certain studies that employ MRI as an input, there is a data type conversion, usually from Dicom to .jpg. This way will certainly alters the information available on the raw data, something we want to avoid when we have to deal with medical images.
- There are still rooms to improve the overall accuracy of CNN based classification of CCs employing MR image analysis, particularly in the case of multiclass classification tasks.

Therefore, the current thesis study aims to fill the gaps customizing the CNN classification system for MR input cervix images.

# Chapter Four

# Materials and Methods

## 4.1. Materials

### 4.1.1. Data Sets

All the MRI dataset were acquired from two PACS (Electronic Picture Archiving and Communication Systems) available at TASH and St. Paul Millennium Hospital where both uses 1.5 T MRI. PACS systems provide economical storage, rapid retrieval of images, access to images acquired with multiple modalities, and simultaneous access at multiple sites. The main goal of having a PACS is to improve operational efficiency while maintaining or improving diagnostic ability [68]. Large medical image data sets within a hospital's PACS combined with advanced high performance parallel computing promises the capacity to accelerate a machine learning technique to more accurately detect clinical imaging findings and diagnose specific diseases. This potential is heightened by the qualifications and characteristics of medical images that make them so ideal for deep learning. Normally, all images come with manually annotated radiologist's reports which are structured to report their findings, impressions, and diagnoses. Often these data are saved in the Electronic Health Record (EHR) system. These reports present an extremely useful resource to apply supervised machine learning algorithms on these images. However, it is often difficult to obtain access to these medical images as they are protected by the Health Insurance Portability and Accountability Act (HIPAA) security rule, which protects a patient's medical records and other personal health information. In the current study, complete consent was obtained from the two hospital sources before data collection.

All the MR images came in raw DICOM (Digital Imaging and Communications in Medicine) format. DICOM is now a global standard that defines how to store, exchange and transmit medical images. Image devices including X-rays, Ultrasound, Microscopy, MRI and CT utilize this standard. DICOM is a slightly unique format as it does not only store the image pixel data (as a special attribute) but also data sets which are made up of attributes (patient name, age, weight). These data sets contain

critical information that must be kept within the file to ensure they are never separated from each other. DICOM files are not limited to one dimensional image. They can represent 2, 3 or 4 dimensional images if the attribute storing the pixel data has multiple frames [69]. The aim of using the published DICOM standard is to ensure that devices made by different companies speak the same language and can, therefore, operate together within the same hospital environment. Radiant DICOM viewer is available for use in visualization of the Dicom images [70]. Additional normal MRI scans were acquired from a freely available online resource.

Figure 4.1 presents two representative T2-weighted axial and sagittal images of a patient treated for stage 4 CC, presented here just as a demo.

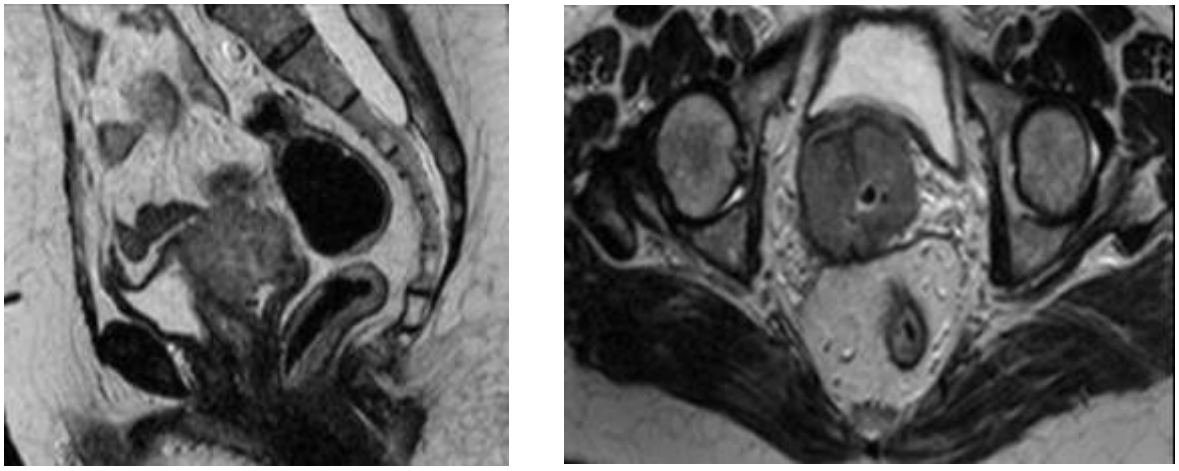


*Figure 4.1: T2-weighted sagittal (left) and axial (right) images.*

### 4.1.2. 3D Slicer

This is a desktop software that could be used to solve advanced image computing challenges with a focus on clinical and biomedical applications. This development platform is used to quickly build and deploy custom solutions for research and commercial products, using free, open-source software. In the current study, 3D Slicer was utilized to extract 2D Dicom image slices from the available 3D MRI stacks [71].

Note that all the different steps of the staging algorithm in the current study was implemented in a Matlab environment (Matlab 2019a). Matlab combines a desktop environment tuned for iterative

analysis and design processes with a programming language that expresses matrix and array mathematics directly.

## 4.2. Methodology

Binary and Multi class classification of CC MR images was considered. Suitable preprocessing techniques were applied on the raw MR images. A schematic block diagram of the proposed method is presented in Fig. 4.2.

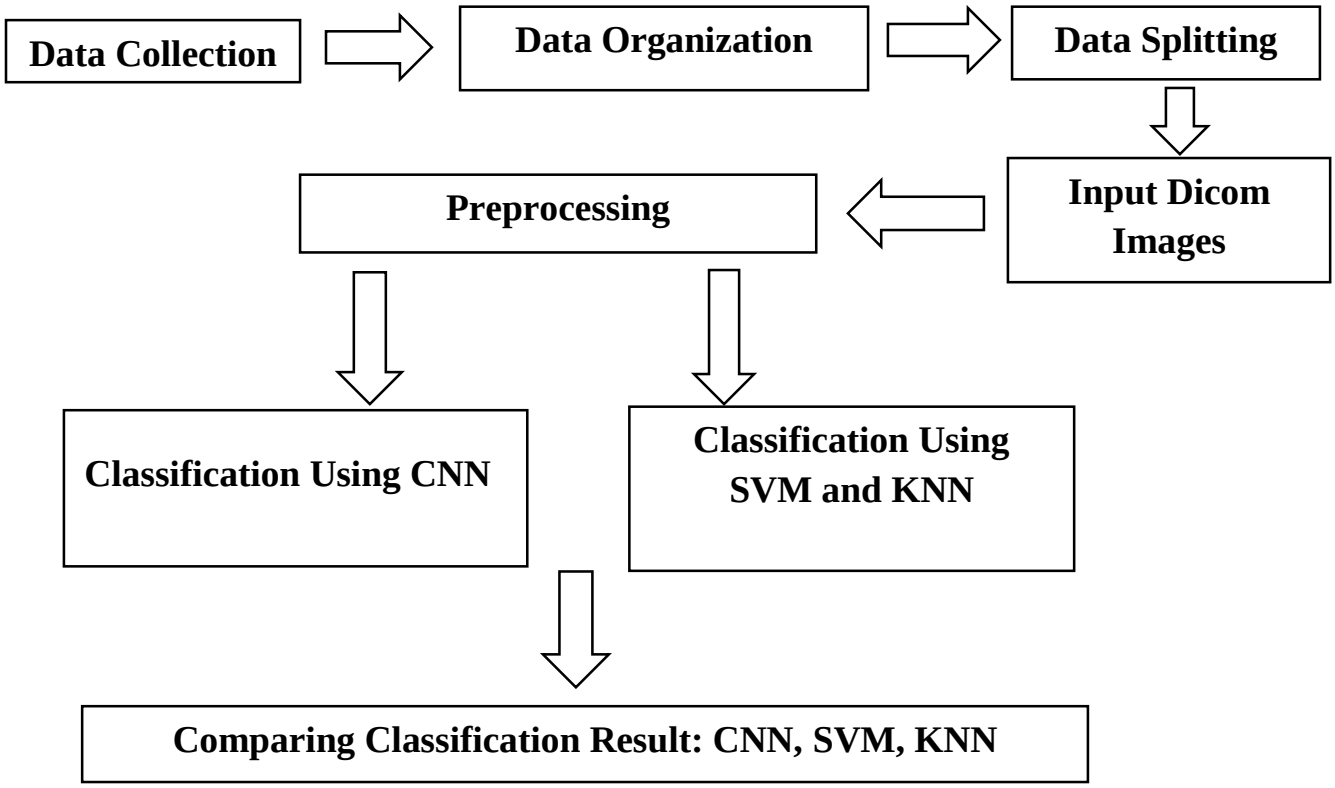



*Figure 4.2: Block Diagram of the Proposed Method.*

### 4.2.1. Step One - Data Collection

Secondary data was acquired from St. Paul and Tikur Anbessa Hospitals in Addis Ababa, Ethiopia (See Appendix D). The collected data were Pelvic MR images stored in the PACS from the year 2017 to 2020. All data came with verified information on the status of the CC stage confirmed by expert oncologists and that has been used as a gold standard in the current study.

### 4.2.2. Step Two - Data Organization

The collected dataset incorporates labeled MR images as T1-weighted, T2-weighted, T2-FLAIR and Diffusion weighted parameters. According to previous researches, Sagittal and Axial T2-weighted images are dominantly used for CC staging procedures. Therefore, in the current research each slice of Axial and Sagittal T2-weighted images were extracted using the 3D slicer software and then organized based on their stage.

### 4.2.3. Step Three – Data Splitting

The data set was divided into training (80% of the total number of patients) and testing (20% of the total number of patients). The data splitting is patient based in order not to use the same person data both in training and testing set [72].

### 4.2.4. Step Four - Pre-Processing

The preprocessing involved two basic steps. First, all the acquired Dicom images were loaded and a 3D matrix (:, :, k) for the images and a column vector of length k for the corresponding labels were generated. Following, image resizing was performed in order to make the input image sizes similar, one requirement to apply the CNN algorithm. The resulting matrices (both the image and label matrices) were then saved as a '.mat file'.

### 4.2.5. Step Five - Customization of CNN Architecture

Details of the CNN architecture and its basic configurations are discussed in the following sections.

**The CNN Architecture**

Multidimensional (md) CNN is used in this project implemented in Matlab. The mdCNN frame work supports all the major network features including dropout, padding, stride, pooling, regularization, momentum, loss function, a classifier and batch normalization as shown in Fig. 4.3 below. The network was preconfigured to run on datasets including MNIST and CIFAR10. The network applied on MNIST offered 99.2% classification accuracy while on CIFAR10 it offered an overall accuracy of 80%. The same framework was also used in another project for classifying vertebra in 3D CT images [73].

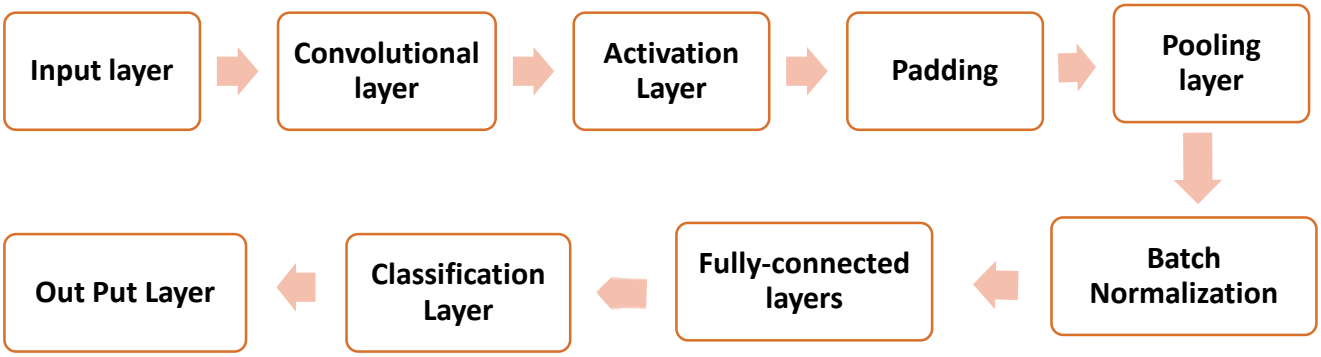



*Figure 4.3: CNN Architecture.*

The mdCNN network structure, training parameters and all other possible configurations are listed in a configuration file. There are 3 groups of network configurations**:** Layer specification**,** Hyper params and Run params. The architecture of the mdCNN is made of multiple layers and each of these layers are briefly discussed below.

**a.** **Input Layer:** is the first layer and specifies the size of a single sample from the dataset. If a large amount of data is fed into a neural network, it will necessitate increased computing and memory requirements.

**b.** **Convolutional Layer**: is the major layer of the network. four convolutional layers were used in binary classification and 2 convolution layer in multi class classification, with the feature maps, padding, pooling and kernel size.

**c.** **Batch Normalization Layer**: normalizes a batch of outputs with 'initGamma' and 'initBeta' as parameters. Gamma and beta are adjusted variables.

**d.** **Fully Connected (FC) Layer:** Every neuron in the FC layer is connected to all outputs of the previous layer. There are two FC layers, the first one has 128 neurons while the second one has five, since the classification class is five and a **'dropOut'** ratio of 0.8 wa**s** employed in the first FC layer.

e. **Classification Layer:** in this case Softmax was used which usually appears after a fully connected layer. This layer performs the Softmax function on the previous layer outputs.
f. **Output Layer:** specifies the loss/cost function for doing back propagation. The two main methods are Mean Square Error (MSE) and Cross Entropy (CrossEnt). MSE is used mainly for regression, and CrossEnt is used for classification. In the current work, CrossEnt was utilized. Function pointers for loss function (derivative of the cost) and cost function need to be provided.

The other important network parameter was the activation function. There are several activation functions one could use, as explained in Chapter 2, including tanh, ReLU, and sigmoid activation functions. In the current study, ReLU activation was employed, as it is now-a-days more popular.

Hyper params are list of parameters which are very important factors to alter the training process. This mainly include number of epochs, train loop count (batch size - the number of samples used in one iteration.), number of test images, image augmentation, learning rate and others. Among the training parameters, the effects of epoch and learning rate were thoroughly investigated in the current study. Epoch is a basic parameter which is defined as the number of passes the machine learning algorithm has made across the entire training dataset. Determining how many epochs a model should run to train is dependent on numerous parameters relating to both the data and the task under consideration. While efforts have been made to turn this process into an algorithm, it is often necessary to have a comprehensive understanding of the data itself. Many models are created with more than one epoch. Even if it is not proved by researchers, there is a relation [74] between the dataset size *D*, number of epochs *E*, number of iterations *I*, and batch size *B* and it is given by:

$$D \times E = I \times B \tag{4.1}$$

### 4.2.6. Training and Testing

This basically involves training of the network with training datasets with their corresponding labels by setting training parameters. This is followed by testing the network with testing datasets and displaying both the confusion matrix and success rate.

To compare and evaluate the efficacy of CNN, two other machine learning algorithms, namely SVM and KNN, were implemented on similar dataset. In CNN, there is no need to extract features separately, but in SVM and KNN, features should be extracted first and the statistical values will be passed to the classifiers. Feature extraction techniques commonly used in medical imaging include intensity histograms, filter-based features, GLCM, scale-invariant feature transform (SIFT) and local binary patterns (LBP). The feature vectors extracted are normally used to train the classification models [75].

### 4.2.7. Performance Measures

Performance measures are extremely powerful tools that indicate the system's capacity to fulfill the intended objective. In the current study, these performance matrices were computed based on the number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN). Based on these measures, a performance matrix is generated that allows to compute the following important performance metrics are calculated as follows:

$$Sensetivity = \frac{TP}{TP+FN} \tag{4.2}$$

$$Specificity = \frac{TN}{FP+TN} \tag{4.3}$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4.4}$$

$$Precision = \frac{TP}{TP+FP} \tag{4.5}$$

$$Recall = \frac{TP}{TP+FN} \tag{4.6}$$

$$F1 - score = \frac{2xPrecision\ x\ Recall}{Precision+Recall} \tag{4.7}$$

### 4.2.8. Graphical User Interface

Graphical User interfaces (GUIs) are the interfaces via which users engage with designs. Matlab APP Designer, an interactive development tool for designing an app layout and programming its behavior, was used to create the user interface in the current study [76]. This interface prompts the user to enter a single sample, after which it employs the feed forward algorithm to select the class with the highest classification probability. At the end, the CC stage of the input sample is displayed.

## 4.3. Employing Support Vector Machine

Basic preprocessing was used for the SVM classification. The preprocessed images were then subjected to feature extraction to gain useful information.

**Steps in Preprocessing**

I. **Cropping***:* The primary goal of this procedure is to pinpoint the area of interest. The input images were all trimmed to a size of 200 x 200 pixels, which essentially shows the important regions (cervix, vagina, bladder, and rectum) on both sagitial and axial T2 Weighted images. The cropping technique calculates the image's center first, then uses the Matlab built-in function imcrop to acquire the desired size.

II. **Thresholding***:* Otsu thresholding is used to segment the images based on its histogram (intensity value distribution). Otsu is a power full technique which helps to binarize the input images. It works by calculating the inter class variance between the foreground and background.

III. **Feature Extraction:** On the resulted binary image, basic region descriptors, namely area, perimeter and centroid were calculate using Matlab built in function 'region props'. Region props measures properties of the image region. Mean and standard deviation of the input images were also calculated. In addition, four basic textural features, namely Energy, Contrast, Entropy, and Homogeneity were extracted from preprocessed images based on the GLCM. As it is indicated on literatures [77 - 80], GLCM is the most prevalent feature extraction technique. It establishes the textural relationship between pixels by executing an operation on the images' second-order statistics.

Nine-features namely Area, Perimeter, Centroid, Mean, Standard Deviation, Energy, Contrast, Entropy, and Homogeneity were finally extracted and concatenated in one vector.

Then the feature vectors were inputted to the SVM classifier employing the Matlab built-in function ‘fitcsvm’. To enhance the classification accuracy, parameter tuning was conducted with and without kfold cross validation.

### 4.4. Employing K-Nearest Neighbor

K-Nearest Neighbor (K-NN) is based on the similarity between the new case and available cases. K-NN algorithm can be used for regression as well as for classification problems but mostly it is used for classification. K-NN performs classification based on the distance calculated between a test sample and the specified samples. There are many matrices that could be employed to calculate the distance between two sets including Manhattan, Canberra, Euclidean and other distance matrices. K-NN is commonly based on the Euclidean distance between the test sample and the specified training samples [81]. In the current study, the KNN model was designed using three built in Matlab functions: ‘fitcknn’ that returns a KNN classification model based on the input variables, ‘crossval’ that performs the loss estimate using cross-validation and ‘kfold’ loss. The KNN model was trained using the same preprocessed datasets that were used both in the SVM models.

# Chapter Five

## Results and Discussion

Raw Dicom MR Images acquired under each of the five classes, including the four CC stages and normal controls, were used as inputs to the proposed CNN based classification scheme following the preprocessing steps. The gathered dataset shows that the data is unbalanced. This is expected and very well agrees with a report made in one of the previous CC researches conducted in Addis Ababa, Ethiopia that showed it is often the case that most of the CC patients admitted to hospitals fall either under stage 2 or stage 4 [17]. Most patients in this unbalanced dataset only had T2 axial or sagittal images, and in some cases, none of the two, which will have an impact on the CNN performance. Only patients with both T2 axial and sagittal sequences were then chosen while the rest were discarded. The data set was then subdivided into training (80% of the total number of patients) and testing (the remaining 20%). After splitting, the T2 axial and sagittal sequence slices were retrieved using 3D slicer software. From a single sagittal and axial volume image, the 3D slicer generates 30-40 2D sagittal images and 40-60 2D axial images. Variation in number of slices might lead to data imbalance. Hence, twenty mid slices were chosen from each sequence.

| Class | Total # of patients | Train (patient) | Test (patient) | Training Axial and Sagittal Images Slices | Testing Axial and Sagittal Images Slices |
|---|---|---|---|---|---|
| Normal | 6 | 5 | 1 | 200 | 40 |
| Stage 1 | 8 | 6 | 2 | 240 | 80 |
| Stage 2 | 10 | 8 | 2 | 320 | 80 |
| Stage 3 | 3 | 2 | 1 | 80 | 40 |
| Stage 4 | 9 | 7 | 2 | 280 | 80 |

*Table 5.1: T2 axial and sagittal image dataset.*

Table 5.1 clearly shows that there is considerable data imbalance which will significantly affect both the training and testing performance. In order to balance the dataset over sampling and one augmentation technique, namely image rotation, were considered. While the rotation was effected at 90, 180 and 270 degrees and applied on the normal and stage 3 datasets. Table 5.2 below presents a summary of the size of the dataset finally considered in the case of Multi class and Binary Data set. Accordingly, the total number of normal image slices considered for further study was 800 while the number of abnormal slices were 1960, combining both axial and sagittal sequences.

| Class | Stage | Train Images | Test Images |
|---|---|---|---|
| Normal | Normal | 800 | 320 |
| Abnormal | I | 240 | 80 |
| | II | 320 | 80 |
| | III | 320 | 80 |
| | IV | 280 | 80 |

*Table 5.2: Dataset after Augmentation.*

## 5.1. Performance of the Proposed CNN Based CC Classification Scheme

The performance of the proposed CNN based classification was tested with three datasets. The first dataset incorporates only normal and abnormal cases where the abnormal dataset were created by merging stage 1 up to stage 4 classes. In the second dataset, normal and two abnormal cases (early stages merging stage 1&2 and an advanced stage merging stage 3&4), the third dataset was composed of each stages of CC. Furthermore, the effect of layer depth and hyper parameter variation such as epoch, kernel size, learning rate and number of filters were investigated.

### 5.1.1. CNN Binary Classification Result

Before obtaining the final result, the effect of different parameters were investigated. Table 5.3 shows the structure of the network layer.

| No. | Number of Layers | Details of Layers |
|---|---|---|
| 1 | 6 | Input, convolution, Fully Connected, Fully Connected, soft max, output |
| 2 | 8 | Input, convolution, convolution, bath norm, Fully Connected, Fully Connected, soft max, output |
| 3 | 11 | Input, convolution, convolution, bath norm, convolution, convolution, bath norm, Fully Connected, Fully Connected, soft max, output |
| 4 | 14 | Input, convolution, convolution, bath norm, convolution, convolution, bath norm, convolution, convolution, bath norm, Fully Connected, Fully Connected, soft max, output |

*Table 5.3: Number of Network Layers.*

Tables 5.4 indicate the effect of layer variation for a given constant epoch 9.

| No | Layer | Overall Accuracy | Normal Accuracy | Abnormal Accuracy |
|---|---|---|---|---|
| 1 | 6 | 78.4% | 72.5% | 84.4% |
| 2 | 8 | 77.3 % | 55.6% | 99.1% |
| 3 | 11 | 85.0 % | 76.3% | 93.8% |
| 4 | 14 | 77% | 58.8% | 95.3% |

*Table 5.4: Effect of Network Layers at Binary classification result.*

The above results clearly show that layer 11 has relatively good accuracy in the case of binary classification. Table 5.5 shows the effect of epoch variation considering layer 11.

| No | Epoch | Overall Accuracy | Normal Accuracy | Abnormal Accuracy |
|---|---|---|---|---|
| 1 | 9 | 85.0 % | 76.3% | 93.8% |
| 2 | 18 | 83.6 % | 80.6% | 86.6% |
| 3 | 30 | 84.2 % | 71.6% | 96.9% |
| 4 | 50 | 84.2% | 71.6% | 96.9% |

*Table 5.5: Binary classification result.*

The other parameter is Kernel size. Kernel is a small mask that moves over the image for feature extraction. Kernel size refers to the width x height of the filter mask. The effect of the most commonly used kernel sizes which are 3x3, 5x5 and 7x7 were investigated and the result is presented in table 5.6. Kernel size of 5x5 gives better accuracy, therefore it is used for the current. Furthermore, the effect of another hyper parameter that influences system performance (learning rate) was also investigated. The current study takes into account a learning rate scheduler with initial and final learning rates. The learning rate is usually set at 0.1 or 0.01. Considering the default and other two values as indicated in table 5.6, learning rate initial value at 0.05 gives better accuracy.

| No | Kernel size | Over all Accuracy |
|---|---|---|
| 1 | 3x3 | 72.5% |
| 2 | 5x5 | 85.0 % |
| 3 | 7x7 | 80.6% |

| No | Learning rate | | Over all Accuracy |
|---|---|---|---|
| | Initial | Final | |
| 1 | 0.5 | 0.0001 | 67% |
| 2 | 0.05 | 0.0001 | 85.0 % |
| 3 | 0.1 | 0.0001 | 83.6% |
| 4 | 0.01 | 0.0001 | 74.5% |

*Table 5.6: Effect of kernel size (left) and learning rate (right) on Binary classification result at a constant Epoch 9 and using 11 layer network architecture.*

In training, the number of filters is also another aspect that influences system performance. The number of filters equals the number of neurons since each neuron performs a distinct convolution on the input, resulting in 'N' feature maps, where 'N' is the number of filters used. Usually, the number of filters is a multiple of two, and in this work, assuming four convolutional layers, eight, twelve, sixteen, and twenty-four filters yields good result in the case of binary classification.

Therefore the final binary classification result using 11 network layer with 5x5 kernel size and 0.05 initial learning rate is presented in fig 5.1 below:

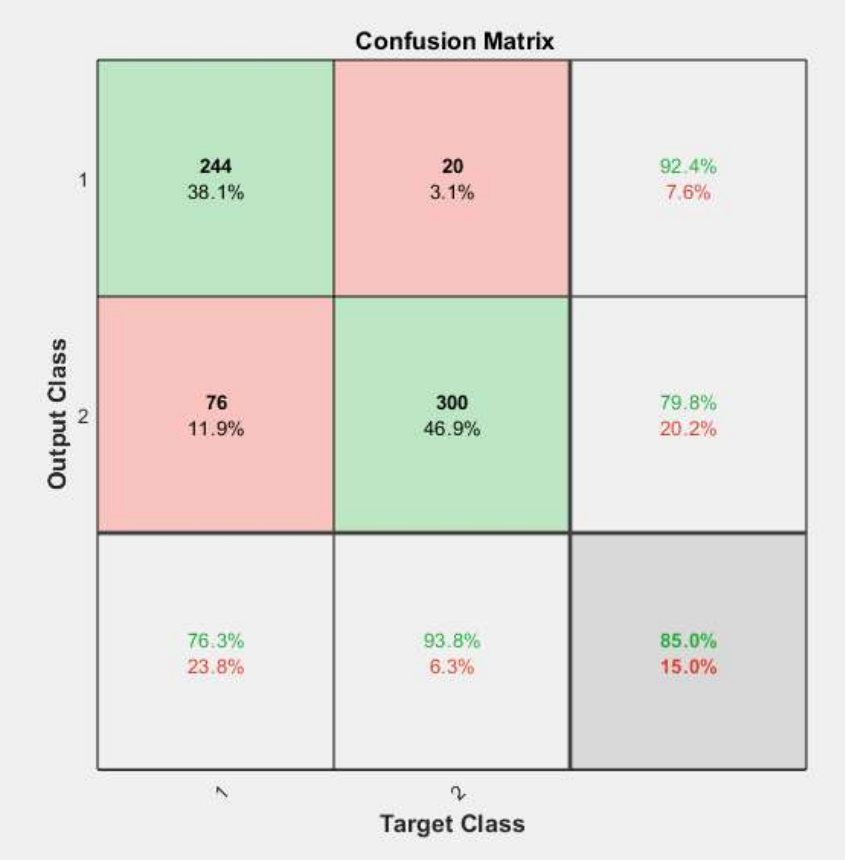



*Figure 5.1: Two Class CNN Classification result @Epoch 9.*

### 5.1.2. CNN Multiclass classification

As table 5.1 indicates, the original dataset is unbalanced and data in each class is very small for training the classifier. As stated before, rotation and over sampling creates new dataset. Under this multi class classification, three class and five class classification were considered. Three class is organized as Normal, Stage 1 & 2 together as an early stage and Stage 3 &4 together as an advanced stages. In the case of five class classification Normal and four each stages of Abnormal cases were considered.

## I. Three Class Classification Result

As it was considered in the case of binary classification, Effect of network layer depth, Epoch, Kernel Size and Learning rate were evaluated. 14 Network layer gives better accuracy than the rest as it can be seen in table 5.7.

| No | Layer | Overall Accuracy | Normal Accuracy | Abnormal Accuracy | |
|---|---|---|---|---|---|
| | | | | Early (Stages 1&2) | Advanced (Stage 3&4) |
| 1 | 6 | 58.3% | 66.3% | 66.9% | 33.8% |
| 2 | 8 | 56.3% | 61.6% | 71.9% | 30% |
| 3 | 11 | 65.3% | 76.9% | 31.3% | 76.3% |
| 4 | 14 | 68.4% | 97.2% | 27.5% | 51.9% |

*Table 5.7: Effect of Number of Network Layers at three class classification result.*

| No | Epoch | Overall Accuracy | Normal Accuracy | Abnormal Accuracy | |
|---|---|---|---|---|---|
| | | | | Early (Stages 1&2) | Advanced (Stage 3&4) |
| 1 | 9 | 68.4% | 97.2% | 27.5% | 51.9% |
| 2 | 18 | 68.8% | 83.4% | 34.4% | 73.8% |
| 3 | 30 | 68.3% | 86.3% | 43.8% | 56.9% |
| 4 | 50 | 66.3% | 85.3% | 40.6% | 53.8% |

*Table 5.8: Three class classification result.*

| No | Kernel size | Over all Accuracy |
|---|---|---|
| 1 | 3x3 | 63% |
| 2 | 5x5 | 68.4% |
| 3 | 7x7 | 65.9% |

| No | Learning rate | | Over all Accuracy |
|---|---|---|---|
| | Initial | Final | |
| 1 | 0.5 | 0.0001 | 53.6% |
| 2 | 0.05 | 0.0001 | 68.4% |
| 3 | 0.1 | 0.0001 | 63.9% |
| 4 | 0.01 | 0.0001 | 59.2% |

*Table 5.9: Effect of kernel size (left) and learning rate (right) on three class classification result at a constant Epoch 9 and using 14 network layer.*

As it is presented in table 5.9. Kernel size, 5X5 and initial learning rate 0.05 were resulted better accuracy similar to the binary classification scenario. Based on these hyper parameters, the final three class classification result is presented in fig 5.2 below.

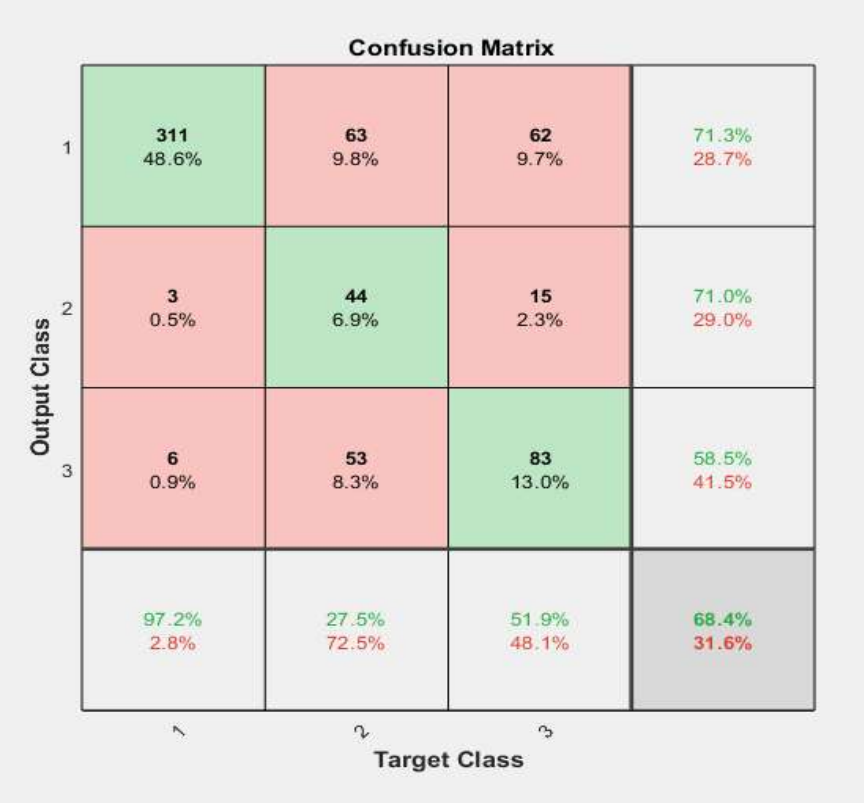



*Figure 5.2: Three Class CNN Classification result @Epoch 9.*

## II. Five Class Classification Result

Similar to the previous classifications, the effect of network layer, epoch number, kernel size and learning rate were evaluated. As it is indicated in table 5.10, 14 layer results a better classification accuracy. Hence it is considered for the final five class classification. However, even if the accuracy increase as the depth of the layer increase, the computation becomes more intensive.

| No | Layer | Overall Accuracy | Normal Accuracy | Abnormal Accuracy | | | |
|---|---|---|---|---|---|---|---|
| | | | | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
| 1 | 6 | 58.1% | 86.3% | 25% | 46.3% | 0 | 48.8% |
| 2 | 8 | 44.2% | 67.2 % | 23.8% | 48.8% | 0 | 12.5% |
| 3 | 11 | 48.3% | 72.5% | 23.8% | 63.7% | 2.5% | 6.3% |
| 4 | 14 | 56.9% | 97.2% | 5% | 60% | 0 | 21.3% |

*Table 5.10: Effect of Number of Network Layers at five class classification result.*

| No | Epoch | Overall Accuracy | Normal Accuracy | Abnormal Accuracy | | | |
|---|---|---|---|---|---|---|---|
| | | | | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
| 1 | 9 | 56.9% | 97.2% | 5% | 60% | 0 | 21.3% |
| 2 | 18 | 50.6% | 77.5% | 2.5% | 60% | 0 | 32.5% |
| 3 | 30 | 49.5% | 75.6% | 18.8% | 47.5% | 0 | 27.5% |
| 4 | 50 | 49.5% | 75.6% | 18.8% | 47.5% | 0 | 27.5% |

*Table 5.11: Five class classification result.*

| No | Kernel size | Over all Accuracy |
|---|---|---|
| 1 | 3x3 | 43.3% |
| 2 | 5x5 | 56.9% |
| 3 | 7x7 | 48.1% |

| No | Learning rate | | Over all Accuracy |
|---|---|---|---|
| | Initial | Final | |
| 1 | 0.5 | 0.0001 | 44.8% |
| 2 | 0.05 | 0.0001 | 56.9% |
| 3 | 0.1 | 0.0001 | 55% |
| 4 | 0.01 | 0.0001 | 50.3% |

*Table 5.12: Effect of kernel size (left) and learning rate (right) on five class classification result at a constant Epoch 9 and using 14 network layer.*

Similar to the other classification results (two class and three class), here also Kernel size 5x5 and 0.05 initial learning rate gives better result as presented in table 5.12 above.

Figure 5.3 below then shows the best five class classification result using 14 network layer with 5x5 kernel size and 0.05 initial learning rate. But in this classification, the result of stage 3 was null. This is due to the small number of patients under stage 3 and even the usage of augmentation did not give the intended increment in classification accuracy.

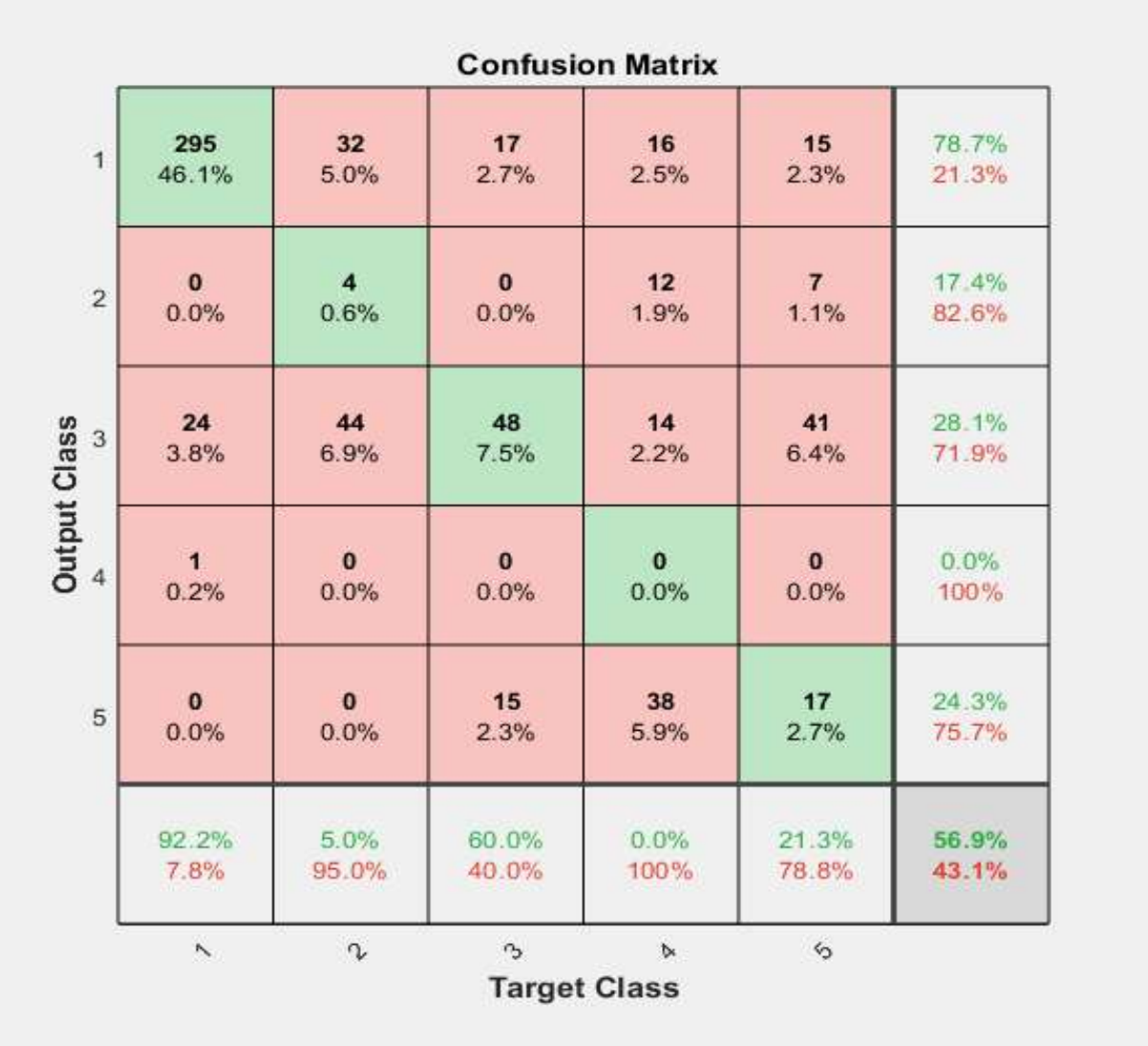



*Figure 5.3: Five Class CNN Classification Result.*

## 5.4. Result of SVM and KNN

SVM and KNN classifier with distinct feature extraction technique was employed on similar dataset used in CNN. Table 5.13 shows the two, three and five class classification results. For SVM classifier in the two class case, linear kernel function and for the multi classification rbf kernel function was used as it is recommended in different sources [82].

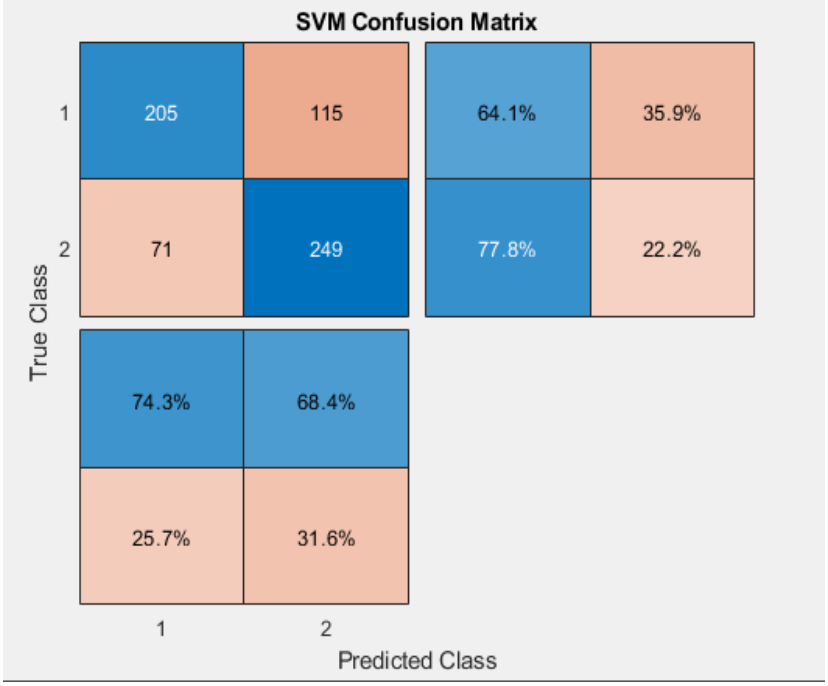
| No | Method | Two class Classification | | |
|---|---|---|---|---|
| | | Over all Accuracy | Normal | Abnormal |
| 1 | SVM | 71 % | 64.1% | 77.8% |
| 2 | KNN | 69% | 65.3% | 72.8% |

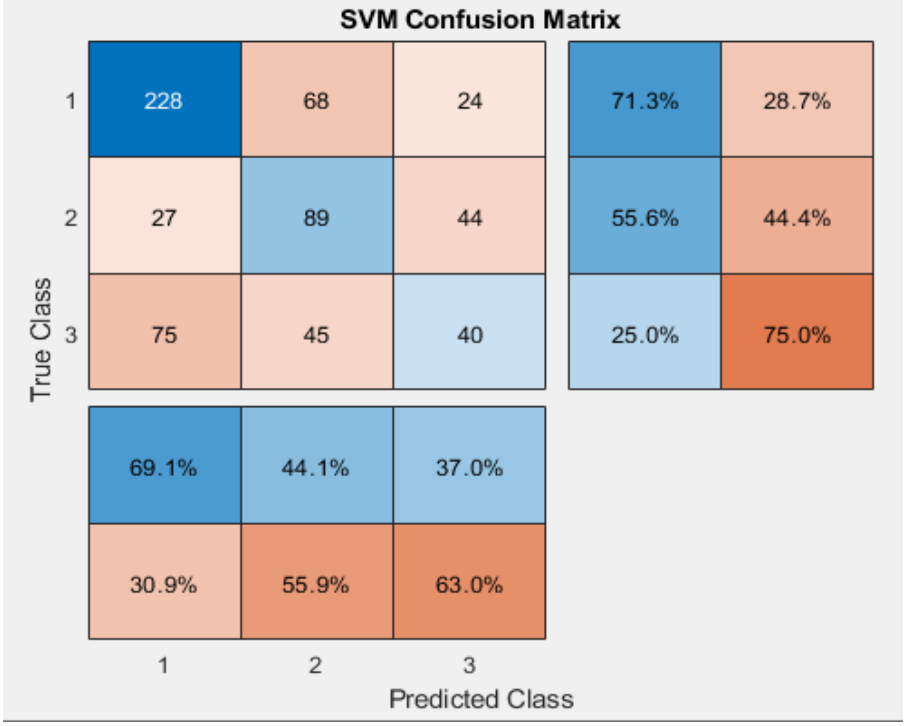
| No | Method | Three class Classification | | | |
|---|---|---|---|---|---|
| | | Over all Accuracy | Normal | Early Stage | Advanced Stage |
| 1 | SVM | 56% | 71.3% | 55.6% | 25% |
| 2 | KNN | 52% | 69.1% | 59.4% | 29.4% |

| No | Method | Five class Classification | | | | | |
|---|---|---|---|---|---|---|---|
| | | Over all Accuracy | Normal | Stage 1 | Stage 2 | Stage 3 | Stage 4 |
| 1 | SVM | 47.3% | 85.6% | 12.5% | 13.8% | 0 | 10% |
| 2 | KNN | 49.8% | 78.1% | 26.3% | 25% | 5% | 30% |

*Table 5.13: SVM and KNN classification result of two class (top), three class (middle) and five class (bottom).*

The confusion matrix for each three classification scenario (two, three and five class) of SVM classifier is shown in fig 5.4 below.


SVM Confusion Matrix
1
2
True Class
205
115
64.1%
35.9%
71
249
77.8%
22.2%
74.3%
68.4%
25.7%
31.6%
1
2
Predicted Class


SVM Confusion Matrix
1
2
3
True Class
228
68
24
71.3%
28.7%
27
89
44
55.6%
44.4%
75
45
40
25.0%
75.0%
69.1%
44.1%
37.0%
30.9%
55.9%
63.0%
1
2
3
Predicted Class

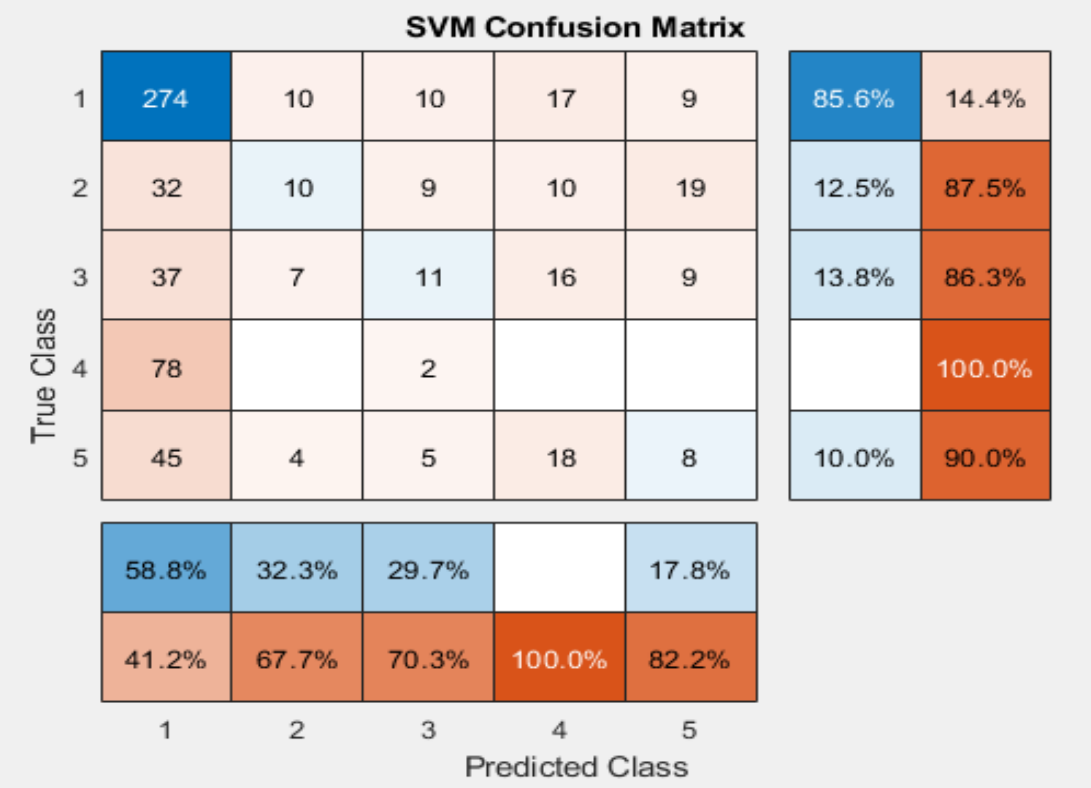
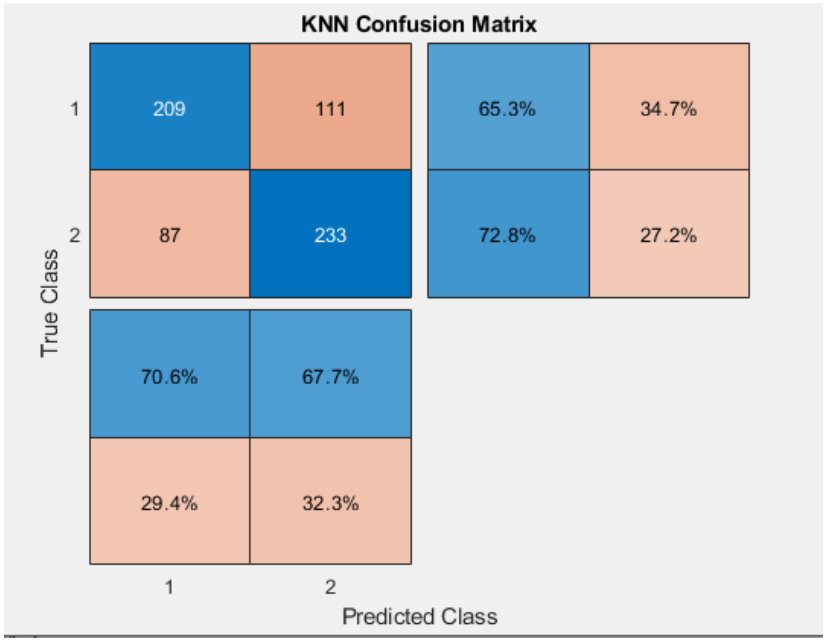
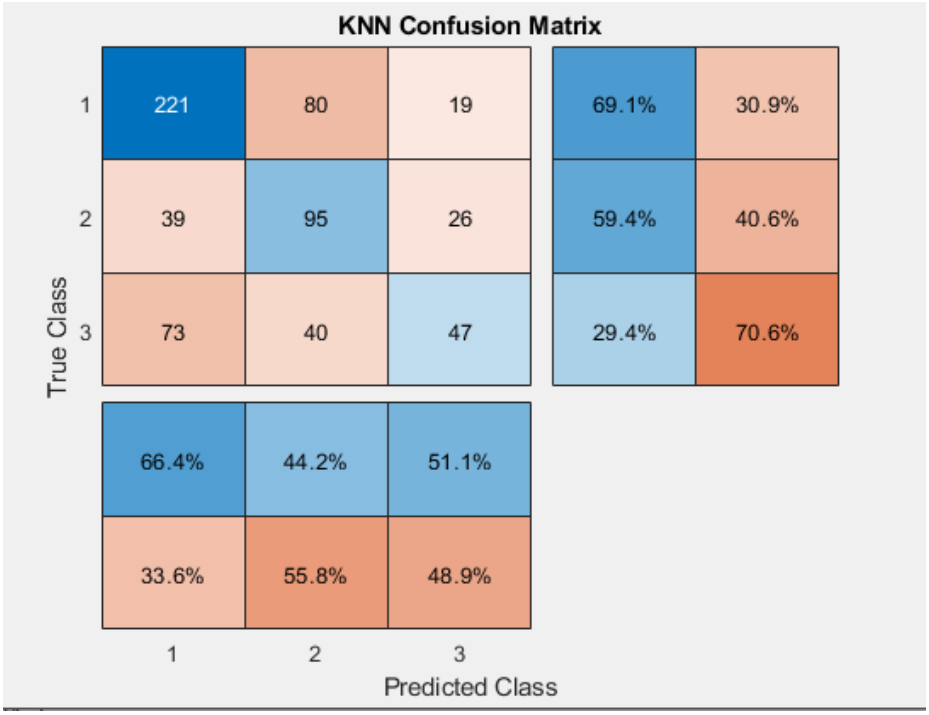



*Figure 5.4: SVM Classification results: Two class (top), Three Class (Middle) and Five Class (bottom).*

The result of KNN classifier for each classification cases is presented in fig 5.5 below.


KNN Confusion Matrix
1
2
209
111
65.3%
34.7%
87
233
72.8%
27.2%
True Class
70.6%
67.7%
29.4%
32.3%
1
2
Predicted Class


KNN Confusion Matrix
1
2
3
221
80
19
69.1%
30.9%
39
95
26
59.4%
40.6%
73
40
47
29.4%
70.6%
True Class
66.4%
44.2%
51.1%
33.6%
55.8%
48.9%
1
2
3
Predicted Class

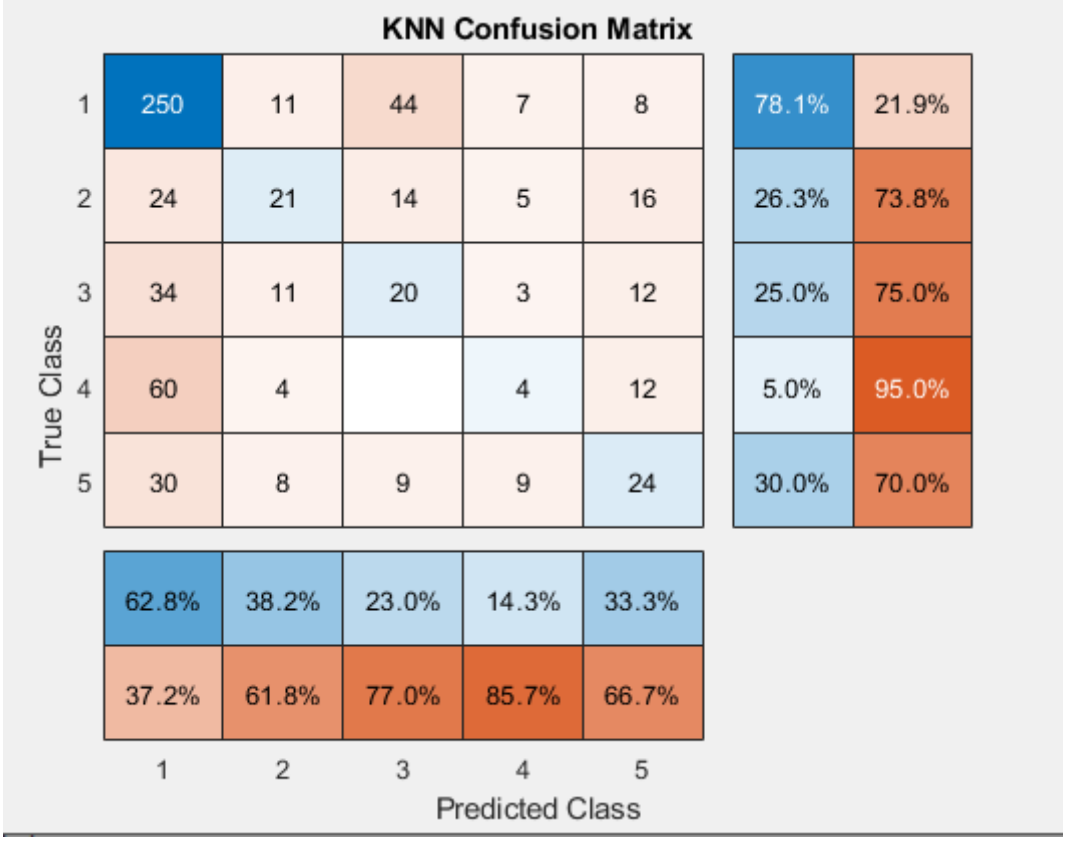



*Figure 5.5: KNN Classification Result: Two class (top), three Class (Middle) and five class (bottom)*

Similar to the case of CNN, five class Classification result of stage three is almost null in both SVM and KNN classifiers. As mentioned above, the reason is due to the use of very small number of training data.

## 5.5. Performance Metrics

In chapter four, the formula of five basic performance metrics were indicated and using those the results in table 5.14 were calculated.

| Metrics | CNN | | |
|---|---|---|---|
| | Two Class | Three Class | Five Class |
| Accuracy | 85% | 68.4% | 56.8% |
| Specificity | 93.8% | 81.6% | 88% |
| Sensitivity | 76.3% | 58.8% | 35.7% |
| Precession | 92.4% | 66.8% | 29.7% |
| F1- Score | 83.6% | 58.9% | 30.7% |

| Metrics | SVM | | |
|---|---|---|---|
| | Two Class | Three Class | Five Class |
| Accuracy | 70.9% | 55.8% | 47.3% |
| Specificity | 77.8% | 76.8% | 82.8% |
| Sensitivity | 64.1% | 50.6% | 24.4% |
| Precession | 74.2% | 50% | 27.7% |
| F1- Score | 68.8% | 49.7% | 23.9% |

| Metrics | KNN | | |
|---|---|---|---|
| | Two Class | Three Class | Five Class |
| Accuracy | 69% | 56.7% | 49.8% |
| Specificity | 72.8% | 76.9% | 84.6% |
| Sensitivity | 65.3% | 52.6% | 32.9% |
| Precession | 70.6% | 53.8% | 34.3% |
| F1- Score | 67.9% | 51.9% | 32.7% |

*Table 5.14: Performance metrics*: CNN (top), SVM (Middle) and KNN (bottom).

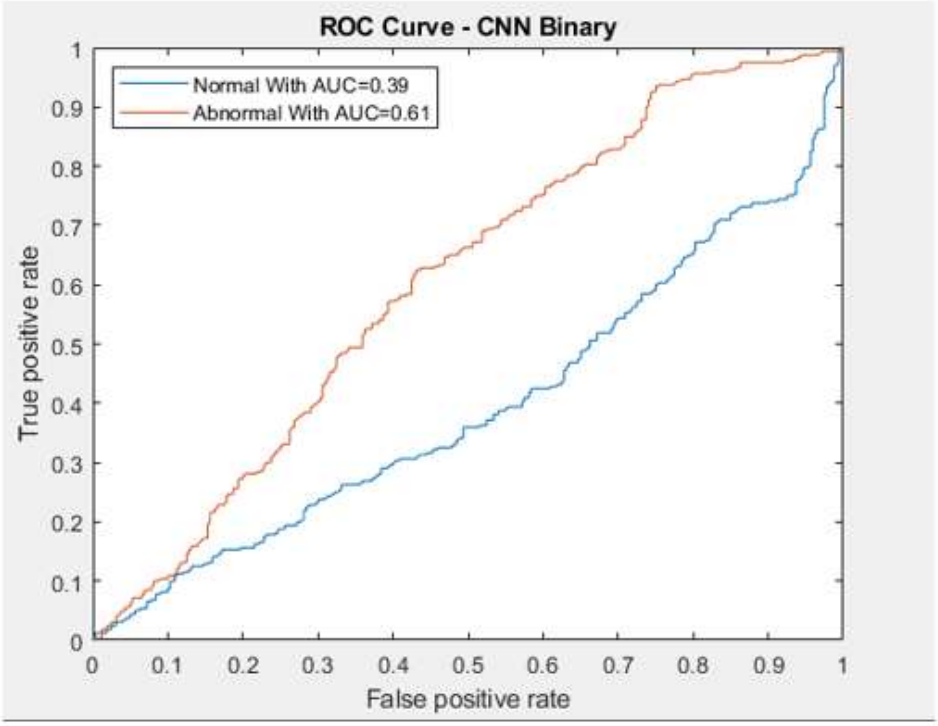
The above result can show that CNN outperform SVM and KNN classifiers. As expected, regarding the computation time, the SVM and KNN classifiers runs much faster than that of the neural network based scheme as the CNN takes more time particularly during training stage.

### 5.6. ROC Analysis

A receiver operator characteristics (ROC) curve is a plot relating the value of true positive rate (TPR) and false positive rate (FPR). It is known as a probability curve since it is plotted based on the classification score probability values of each class. The area under this ROC curve is commonly calculated using trapezoidal rule. This area under curve (AUC), indicates the degree of the classifier to distinguish between classes [83].

#### 5.6.1. Two Class Classification Result

Figure 5.6 indicates higher TPR for abnormal class with higher area under curve value. Comparing the three classifier models (CNN, SVM and KNN), CNN shows a higher degree of separability between normal and abnormal classes than the other two classifiers.



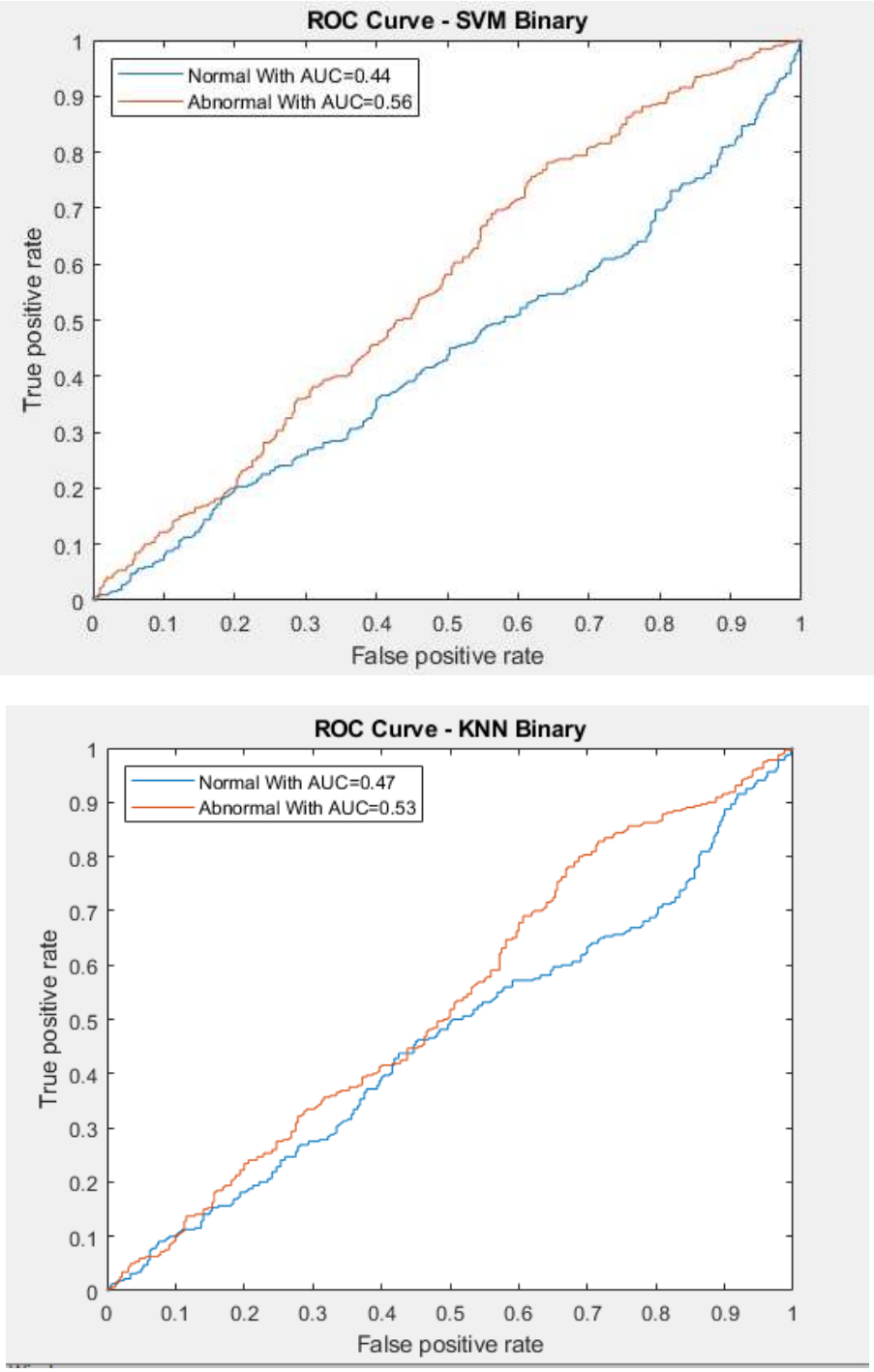
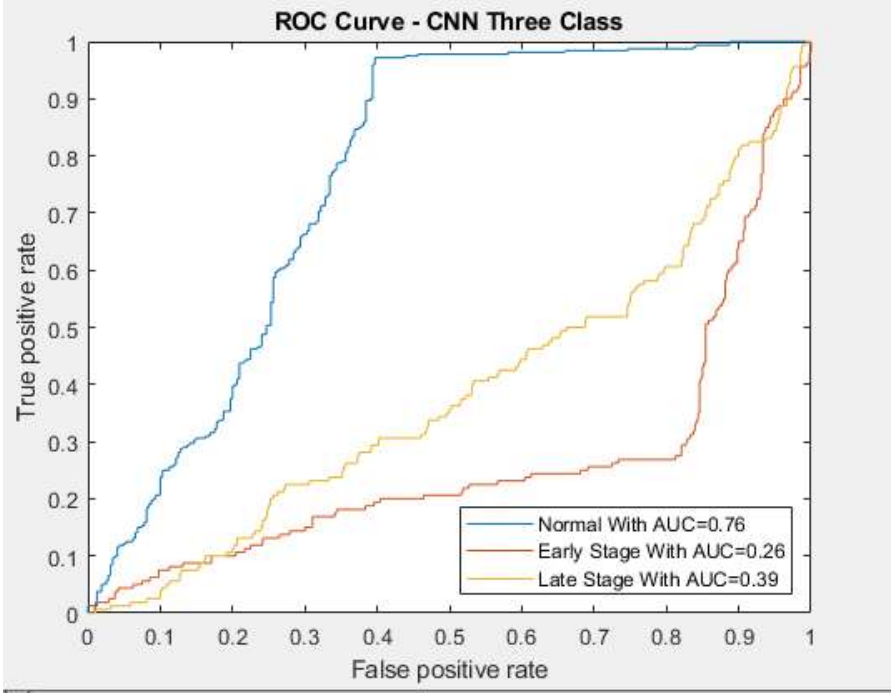



*Figure 5.6: Two Class ROC Analysis: CNN (top), SVM (Middle) and KNN (bottom).*

### 5.6.2. Three class Classification

Figure 5.7 indicates higher TPR for normal class with higher area under curve value. Comparing the three classifier models (CNN, SVM and KNN), CNN shows a higher degree of separability between normal, early stage and advanced stages than the other two classifiers.



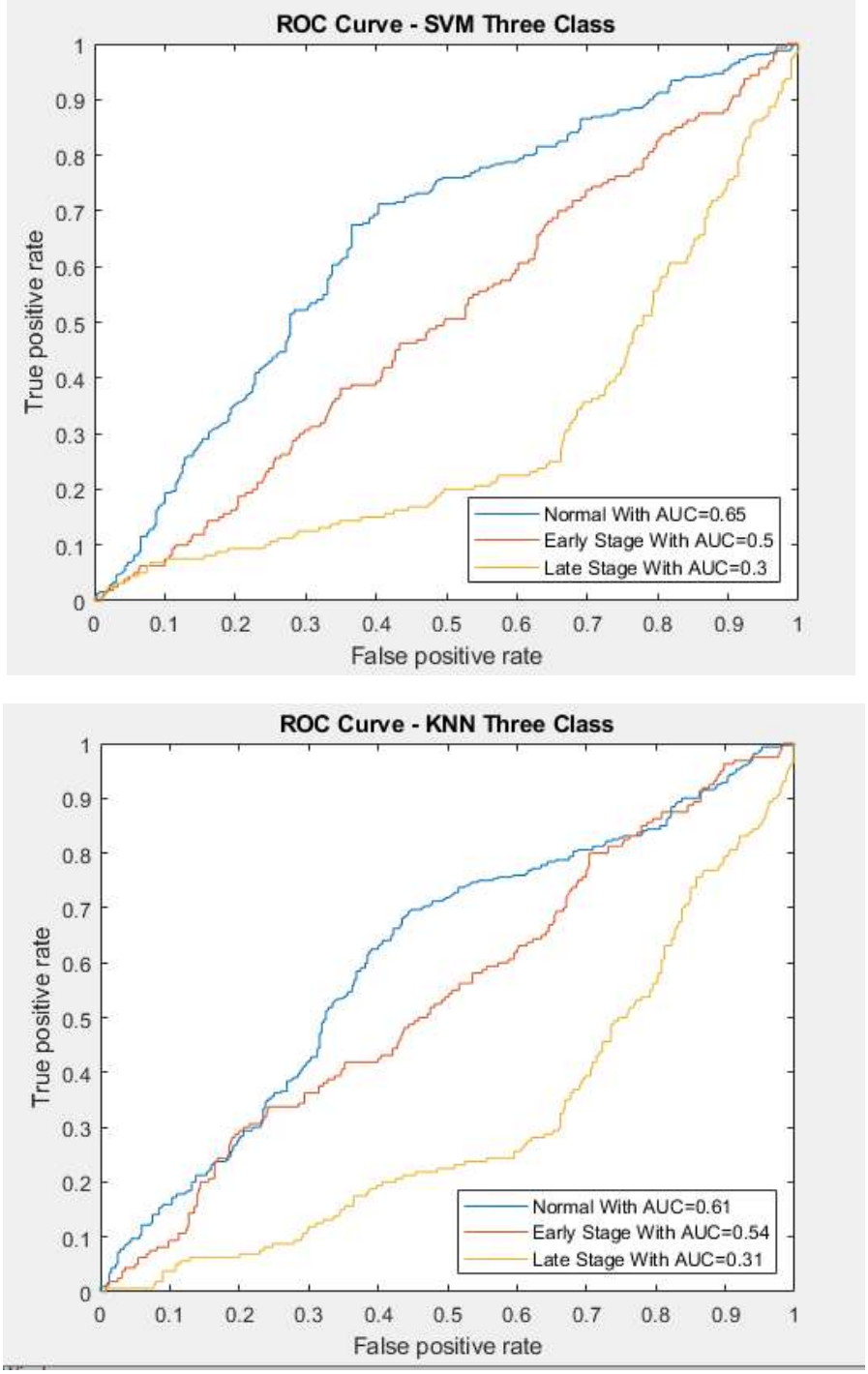
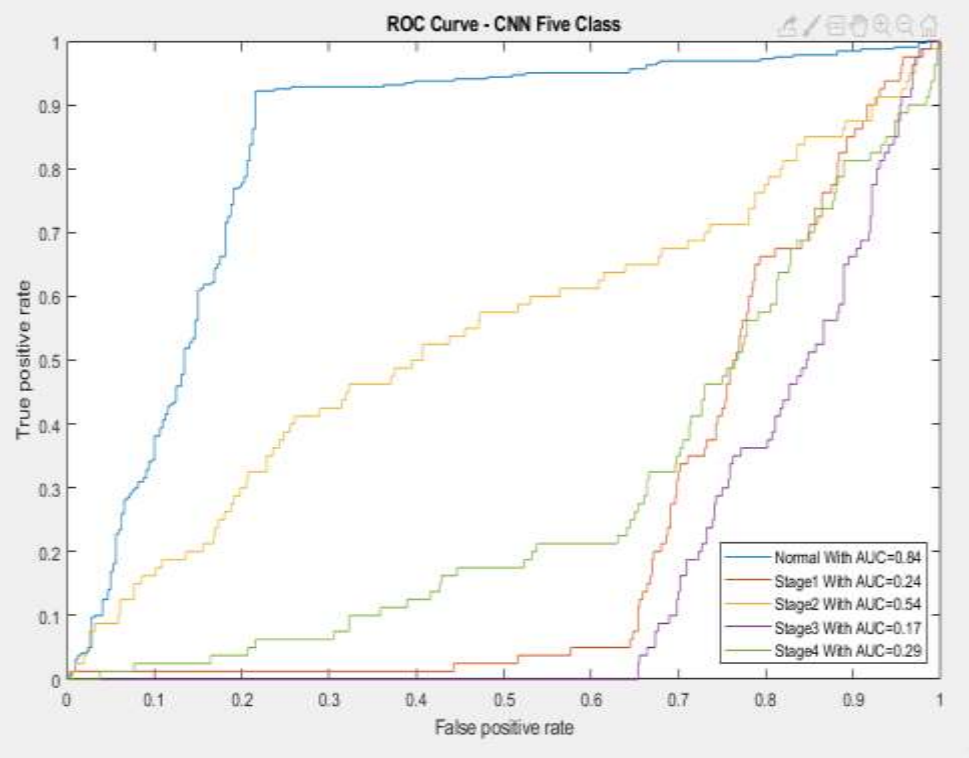
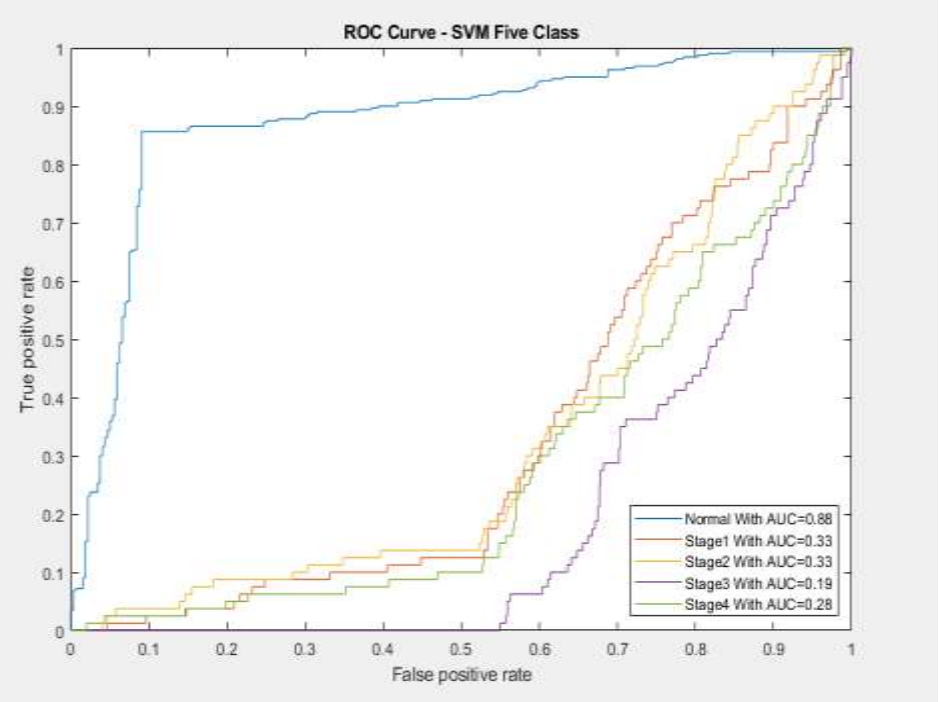



*Figure 5.7: Three Class ROC Analysis: CNN (top), SVM (Middle) and KNN (bottom).*

### 5.6.3. Five Class ROC

Figure 5.8 indicates higher TPR for normal class with higher area under curve value. Comparing the three classifier models (CNN, SVM and KNN), CNN shows a higher degree of separability between normal, stage 1 – 4 classes than the other two classifiers. In this plot, the separability degree between abnormal stages (stage 1 - 4) and the AUC for stage 3 is very low. This is due to the use of small training data to train the classifiers.





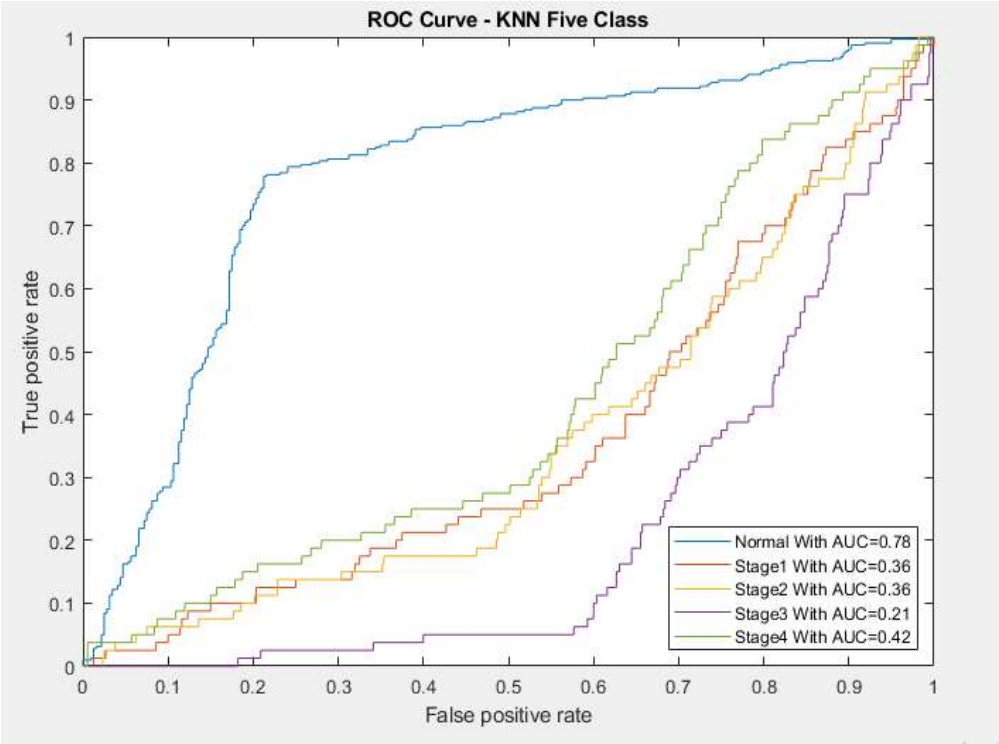




*Figure 5.8: Five Class ROC Analysis: CNN (top), SVM (Middle) and KNN (bottom)*

# Chapter Six

# Conclusion and Recommendations

## 6.1. Conclusion

Cervical cancer is the cause for many women deaths worldwide, with the highest proportion in developing countries. It needs effective diagnosis and treatment in order to minimize the death rate. The proposed system in the current study has a potential of supporting CC staging procedures performed in the clinics thereby easing the time consuming and subjective manual procedures. In rural areas where physicians are scarce, the system can support a decision, benefiting both the patient as well as the care givers. The main architecture of the proposed system is based on two-dimensional convolutional neural network. Pelvic MR images in their standard format (DICOM) were used as inputs to the classifier following application of basic preprocessing techniques. Both binary and multi class classification were considered, In the case of binary classification, Eleven Network Layer 11 with 5x5 kernel and 0.05 initial learning rate achieved a better accuracy which is 85 %. In multi class scenario, three class and five class classification were considered. In multi class classification, fourteen network layer with 5x5 kernel and 0.05 initial learning rate achieved a better accuracy. In this classification, three class and five classes were considered. For the three class the best CNN accuracy was 68.8% and for five class 56.9 %. The output of the CNN scheme was also compared with other machine learning algorithms, namely SVM and KNN where both of them were designed using region descriptors and GLCM as a feature extractor. The results clearly showed that CNN outperforms these two classifiers showing its great promise to help physicians as a decision support system.

The system will still need to be trained and validated on a larger data set to check its robustness. In principle, the methodology explained in the current study for use in automatic staging on CCs based on MR image analysis could still be applicable when one has to use other imaging modalities other than MRI, such as CT and US. The model could be retrained to fix optimal model parameters to be used in analyzing non MR images.

Compared to the other two classifiers discussed in the current study, the CNN based scheme was computationally more intensive. Researches to work around such and related issues would require further studies.

## 6.2. Recommendations

a. It is highly recommended to check the performance of the proposed scheme on larger data sets.
b. The issue of data imbalance is still outstanding. Ideally the CC staging should be carried out having a balanced data size among the different classes/stages.
c. An investigation could be carried out on ways of utilizing the capabilities of the proposed CC diagnosis scheme in supporting telemedicine applications in dealing with issues that are pertinent to CC.
d. Estimation of size and side wall invasion as well as total segmentation of cervical tumors could be very interesting, which could have significant effects during treatments.

# Reference


1. Jordan J, Singer A, Jones H, Shafi M, *The cervix*, 2009 Apr 8.
2. Cervix Image, [Online] Available:https://teachmeanatomy.info/pelvis/female-reproductive-tract/cervix/, Accessed on 2020 Dec 12.
3. Waggoner SE, *Cervical cancer*, The Lancet, 2003 Jun 28;361(9376):2217-25
4. Pelvic MRI Images , [Online] Available: https://www.sciencephoto.com/media/693982/view
5. Fielding J, Ming HC, *Magnetic Resonance Imaging of the Pelvic Floor in Biomechanics of the female pelvic floor*, 2016 Jan 1 (pp. 279-289), Academic Press.
6. FIGO Committee Report, *Revised FIGO staging for carcinoma of the cervix uteri*, 17 January 2019. [Online] Available: https://doi.org/10.1002/ijgo.12749. Accessed on September 2019
7. Cohen PA, Jhingran A, Oaknin A, Denny L. *Cervical cancer*. The Lancet. 2019 Jan 12;393(10167):169-82.
8. American Cancer Society, *Risk Factors for Cervical Cancer,* cancer.org | 1.800.227.2345, [Online] Available: https://www.cancer.org/cancer/cervical-cancer/causes-risks-prevention/risk-factors.html, Accessed on 2020 Oct 26.
9. Anorlu RI, *Cervical cancer: the sub-Saharan African perspective. Reproductive health matter*, 2008 Jan 1;16(32):41-9
10. [Online] Available : https://www.worldbank.org/en/country/ethiopia/overview
11. Bruni LB .et al, ICO/IARC information centre on HPV and cancer (HPV information centre), *Human papillomavirus and related diseases in Ethiopia, Summary report*, 22 October 2021.
12. Woldeamanuel YW, Girma B, Teklu AM. *Cancer in Ethiopia*. The Lancet. Oncology. 2013 Apr 1;14 (4):289-90.
13. Getahun F, Mazengia F, Abuhay M, Birhanu Z, *Comprehensive knowledge about cervical cancer is low among women in Northwest Ethiopia*. BMC cancer. 2013 Dec;13(1):1-7
14. Hailu A, Mariam DH, *Patient side cost and its predictors for cervical cancer in Ethiopia: a cross sectional hospital based study*, BMC cancer. 2013 Dec;13(1):1-8
15. Kebede W, Kebede K, *Psychosocial experiences and needs of women diagnosed with cervical cancer in Ethiopia. International Social Work*. 2017 Nov; 60 (6):1632-46.

16. Begoihn M et al., *Cervical cancer in Ethiopia–predictors of advanced stage and prolonged time to diagnosis. Infectious agents and cancer*. 2019 Dec; 14 (1):1-7.
17. Dereje N, Gebremariam A, Addissie A, *et al*., *Factors associated with advanced stage at diagnosis of cervical cancer in Addis Ababa, Ethiopia: a population-based study, BMJ Open* 2020;10:e040645. doi: 10.1136/bmjopen-2020-040645
18. Moelle U et al., *Cervical cancer in Ethiopia: The effect of adherence to radiotherapy on survival*, The oncologist. 2018 Sep;23(9):1024
19. Kantelhardt EJ et al., *Cervical cancer in Ethiopia: survival of 1,059 patients who received oncologic therapy*. The oncologist. 2014 Jul;19(7):727
20. FIGO Committee Report, *Revised FIGO staging for carcinoma of the cervix uteri*, 17January 2019. [Online] Available: https://doi.org/10.1002/ijgo.12749. Accessed on September 2019.
21. Woldeamanuel YW, Girma B, Teklu AM, *Cancer in Ethiopia*, Lancet Oncology, 2013; 14(4):289–90. [Online] Available: http://dx.doi.org/10.1016/S1470-2045(12)70399-6, Accessed on 2019 Nov 26.
22. Haileselassie W, Mulugeta T et al., *The situation of cancer treatment in Ethiopia: challenges and opportunities. Journal of cancer prevention*. 2019 Mar; 24(1):33. [Online] Available : https://www.cancer.net/cancer-types/cervical-cancer/diagnosis
23. [Online] Available : https://www.cancer.net/cancer-types/cervical-cancer/diagnosis
24. Follen M, Levenback CF, Iyer RB et al., *Imaging in cervical cancer. Cancer*, Interdisciplinary International Journal of the American Cancer Society, 2003 Nov 1; 98(S9):2028-38.
25. Kay, H, Spritzer, C, Glob. libr. women's med.,(ISSN: 1756-2228) 2008; DOI 10.3843/GLOWM.10067
26. P. Sriramakrishnan, T. Kalaiselvi, M et al., *A Role of Medical Imaging Techniques in Human Brain Tumor Treatment*, International Journal of Recent Technology and Engineering International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4S2, 2019 Dec.
27. Broadhouse K, *The Physics of MRI and How We Use It to Reveal the Mysteries of the Mind. Front*, Young Minds. (2019) 7:23. doi: 10.3389/frym.2019.00023

28. Jones, J., Baba, Y. *T1 Weighted Image*. Reference article, Radiopaedia.org. (accessed on 14 Sep 2021) https://radiopaedia.org/articles/5852
29. David C Preston, MD, *Magnetic Resonance Imaging (MRI) of the Brain and Spine: Basics*, Revised 11/30/06 Copyrighted 2006
30. Kurata Y et al., *Automatic Segmentation of the uterus on MRI using convolutional neural network*, Computers in biology and medicine. 2019 Nov 1; 114:103438.
31. Camisão CC et al., *Magnetic resonance imaging in the staging of cervical cancer. Radiologia Brasileira.* 2007; 40:207-15.
32. LudwikaTarłowskaetal., *Comparison of the FIGO and TNM staging systems for uterine cervix cancer based on classification of 6193 cases*, Gynecologic Oncology, Volume 4, Issue 3, 1976, Pages 270-277, ISSN 0090-8258, https://doi.org/10.1016/0090-8258(76)90032 (https://www.sciencedirect.com/science/article/pii/0090825876900329)
33. Healthella Staff, *Cervical Cancer Stages; FIGO, TNM Staging & Metastatic Spread*, June 16 2019, [Online] Available: https://healthella.com/cervical-cancer-stages/. Accessed on 2019 Sep 3.
34. J ClinPathol, *Which staging system to use for gynaecological cancers: a survey with recommendations for practice in the UK*, 2010 Sep; 63(9):768-70. doi: 10.1136/jcp.2010.080978. Epub 2010 Aug 9
35. Otero-garcía MM, Mesa-álvarez A et al., Role of MRI in staging and follow-up of endometrial and cervical cancer : pitfalls and mimickers, Insights into Imaging, 2019; 10(19)
36. Singh Y, Srivastava D, Chandranand PS, Singh S, *Algorithms for screening of Cervical Cancer : A chronological review*, 2018 Nov (Unpublished).
37. Ongsulee P, *Artificial intelligence, machine learning and deep learning*, In2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE) 2017 Nov 22 (pp. 1-6). IEEE.
38. [Online] Available :https://pub.towardsai.net/building-neural-networks-from-scratch-with-python-code-and-math-in-detail-i-536fae5d7bbf
39. O'Shea K, Nash R, *An introduction to convolutional neural networks*, arXiv preprint arXiv: 1511.08458. 2015 Nov 26

40. Li Q, Cai W et al., *Medical image classification with convolutional neural network. In2014 13th international conference on control automation robotics & vision* (ICARCV) 2014 Dec 10 (pp. 844-848). IEEE. .
41. Buda M, Maki A, Mazurowski MA, *A systematic study of the class imbalance problem in convolutional neural networks*, Neural Networks. 2018 Oct 1; 106:249-59.
42. Li Q, Cai W et al., *Medical image classification with convolutional neural network*. In2014 13th international conference on control automation robotics & vision (ICARCV) 2014 Dec 10 (pp. 844-848). IEEE.
43. Anwar SM, Majid M et al., *Medical image analysis using convolutional neural networks*: *a review*. Journal of medical systems. 2018 Nov; 42(11):1-3.
44. Yamashita R, Nishio M, Do RK, Togashi K. *Convolutional neural networks: an overview and application in radiology*. Insights into imaging. 2018 Aug; 9(4):611-29.
45. Ruder S., *An overview of gradient descent optimization algorithms*, arXiv preprint arXiv: 1609.04747. 2016 Sep 15
46. Sarwar A et al., *Performance Evaluation of Machine Learning Techniques for Screening of Cervical Cancer*, in Proceedings of the 2nd International Conference on Computing for Sustainable Global Development, IEEE, 11-13 March 2015, New Delhi, India.
47. William W, Ware A, Basaza-ejiri AH, Obungoloch J, *A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images*, Computer Methods Programs Biomed, 2018;164:15–22. [Online] Available: https://doi.org/10.1016/j.cmpb.2018.05.034. Accessed on November 2019.
48. Bethanney J, Umashankar G, Sindu D, Shelcy M, Nancy S. Classification of Cervical Cancer from MRI Images using Multiclass SVM Classifier. International Journal of Engineering & Technology. 2018;7(2):1-5.
49. Aabha S, Prof. Dr. Ashwini B. *Classification of Mr Images of Cervical Cancer Using SVM and ANN*, Volume : 4 | Issue : 5 | May 2015 • ISSN No 2277 - 8179
50. Pabitra M & Sushmita M, *Staging of Cervical Cancer with Soft Computing*, *IEEE* Transactions on Biomedical Engineering, July 2000; 47(7): 934 – 940.

51. Soumya MK, Sneha K, Arunvinodh C, *Cervical Cancer Detection and Classification using Texture Analysis*, Biomedical & Pharmacology Journal, 2016;9(2):663–71.
52. Mohanaiah P, Sathyanarayana P, GuruKumar L., *Image texture feature extraction using GLCM approach*, International journal of scientific and research publications, 2013 May; 3(5):1-5.
53. Jain S, *Brain cancer classification using GLCM based feature extraction in artificial neural network*, International Journal of Computer Science & Engineering Technology. 2013 Jul 7; 4(7):966-70.
54. Yu H, Kim S, *SVM Tutorial-Classification, Regression and Ranking. Handbook of Natural computing*, 2012; 1:479-506.
55. Yadav SS, Jadhav SM. *Deep convolutional neural network based medical image classification for disease diagnosis*. Journal of Big Data. 2019 Dec; 6(1):1-8.
56. Cernazanu-Glavan C, Holban S. *Segmentation of bone structure in X-ray images using convolutional neural network*. Adv. Electr. Comput. Eng. 2013 Jan 1; 13(1):87-94.
57. Schwendicke F, Golla T, Dreher M, Krois J. *Convolutional neural networks for dental image diagnostics: A scoping review*. Journal of dentistry. 2019 Dec 1; 91:103226.]
58. Alakwaa W, Nassef M, Badr A. *Lung cancer detection and classification with 3D convolutional neural network (3D-CNN)*. Lung Cancer. 2017;8(8):409
59. Charan S, Khan MJ, Khurshid K. *Breast cancer detection in mammograms using convolutional neural network.* In2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) 2018 Mar 3 (pp. 1-5). IEEE.
60. Hashemzehi R, Mahdavi SJ, Kheirabadi M, Kamel SR. *Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE*. Biocybernetics and Biomedical Engineering. 2020 Jul 1; 40(3):1225-32.
61. Zhang X & Zhao S, *Cervical image classification based on image segmentation preprocessing and a CapsNet network model*, Int J Imaging Syst Technol, August 2018;19–28. [Online] Available: wileyonlinelibrary.com/journal/ima. Accessed on November 2019.
62. Nguyen H & Leavitt T, *CS 231N Final Project Report : Cervical Cancer Screening*, 2017. [Online] Avaialble: https:// www.semanticscholars.org/paper. Accessed on 2019 Sept 12.

63. Alyafeai Z, Ghouti L. *A fully-automated deep learning pipeline for cervical cancer classification.* Expert Systems with Applications. 2020 Mar 1; 141:112951.
64. Sitaram S, Dessai A. *Classification of Cervical MR Images Using ResNet101*., 2019
65. Ghoneim A, Muhammad G, Hossain MS, *Cervical cancer classification using convolutional neural networks and extreme learning machines*, Future Generation Computer System. Semptember 2019;102:643–9. [Online] Available: https://doi.org/10.1016/j.future.2019.09.015. Accessed on 2019 Dec 16.
66. Bora K, Chowdhury M, Mahanta LB, Kundu MK, Das AK. *Pap smear image classification using convolutional neural network*. In Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing 2016 Dec 18 (pp. 1-8).
67. Rehman AU, Ali N, Taj I, Sajid M, Karimov KS. *An Automatic Mass Screening System for Cervical Cancer Detection Based on Convolutional Neural Network. Mathematical Problems in Engineering.* 2020 Oct 28; 2020.
68. Choplin RH, Boehme 2nd JM, Maynard CD. *Picture archiving and communication systems: an overview*. Radiographics. 1992 Jan; 12(1):127-9.
69. [Online] Available: https://radiopaedia.org/articles/digital-imaging-and-communications-in-medicine-dicom, Accessed on 2021 Sep 19.
70. Radiant Dicom viewer, [Online] Available: https://www.radiantviewer.com/, Accessed on 2020 Oct 5.
71. 3D slicer, [Online] Available: https://www.slicer.org/ , Accessed on 2019 June 24.
72. Golan R, Jacob C, Denzinger J. *Lung nodule detection in CT images using deep convolutional neural networks*. In2016 International Joint Conference on Neural Networks (IJCNN) 2016 Jul 24 (pp. 243-250). IEEE.
73. Hagay garty , mdCNN, [online] Available :https://github.com/hagaygarty/mdCNN, Accessed on 2020 Nov.
74. Daniel J. et al., *Epoch (machine learning),* [online] Available: https://radiopaedia.org/articles/epoch-machine-learning, Accessed on 2021 May 10.
75. Mohanaiah P, Sathyanarayana P, GuruKumar L. *Image texture feature extraction using GLCM approach*. International journal of scientific and research publications. 2013 May; 3(5):1-5.

76. [Online] Available: https://www.mathworks.com/help/matlab/app-designer.html, Accessed on 2021 May.
77. Öztürk Ş, Akdemir B. *Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA*. Procedia computer science. 2018 Jan 1; 132:40-6.
78. Kumar D. *Feature extraction and selection of kidney ultrasound images using GLCM and PCA*. Procedia Computer Science. 2020 Jan 1; 167:1722-31.
79. Htay TT, Maung SS. *Early stage breast cancer detection system using glcm feature extraction and k-nearest neighbor (k-NN) on mammography image*. In2018 18th International Symposium on Communications and Information Technologies (ISCIT) 2018 Sep 26 (pp. 171-175). IEEE.
80. Sairamya NJ, Susmitha L et al., *Hybrid Approach for Classification of Electroencephalographic Signals Using Time–Frequency Images With Wavelets and Texture Features.* In Intelligent Data Analysis for Biomedical Applications 2019 Jan 1 (pp. 253-273). Academic Press.
81. [Online] Available: Peterson LE. K-nearest neighbor. Scholarpedia. 2009 Feb 21;4(2):1883., Accessed on 2021 Jul.
82. [Online] Available: https://www.mathworks.com/help/stats/fitcsvm.html, Accessed on 2022 Jan.
83. [Online] Available:https://en.wikipedia.org/wiki/Receiver_operating_characteristic,Accessed on 2022 May.

# Appendix A

### FIGO Stages for Cervical Cancer

*Stage I:* The carcinoma is strictly confined to the cervix uteri (extension to the corpus should be disregarded).

- **IA:** Invasive carcinoma that can be diagnosed only by microscopy, with maximum depth of invasion <5 mm
    - ✓ **IA1:** Measured stromal invasion <3 mm in depth
    - ✓ **IA2:** Measured stromal invasion ≥3 mm and <5 mm in depth
- **IB:** Invasive carcinoma with measured deepest invasion ≥5 mm (greater than stage IA), lesion limited to the cervix uteri
    - ✓ **IB1:** Invasive carcinoma ≥5 mm depth of stromal invasion and <2 cm in greatest dimension
    - ✓ **IB2:** Invasive carcinoma ≥2 cm and <4 cm in greatest dimension
    - ✓ **IB3:** Invasive carcinoma ≥4 cm in greatest dimension

*Stage II:* The carcinoma invades beyond the uterus, but has not extended onto the lower third of the vagina or to the pelvic wall.

- **IIA:** Involvement limited to the upper two-thirds of the vagina without parametrial involvement
    - ✓ **IIA1:** Invasive carcinoma <4 cm in greatest dimension
    - ✓ **IIA2:** Invasive carcinoma ≥4 cm in greatest dimension
- **IIB:** With parametrial involvement but not up to the pelvic wall

*Stage III:* The carcinoma involves the lower third of the vagina and/or extends to the pelvic wall and/or causes hydronephrosis or non-functioning kidney and/or involves pelvic and/or paraaortic lymph nodes.

- **IIIA:** Carcinoma involves the lower third of the vagina, with no extension to the pelvic wall
- **IIIB:** Extension to the pelvic wall and/or hydronephrosis or non-functioning kidney (unless known to be due to another cause)

- **IIIC:** Involvement of pelvic and/or paraaortic lymph nodes, irrespective of tumor size and extent (with r and p notations)
    - ✓ **IIIC1:** Pelvic lymph node metastasis only
    - ✓ **IIIC2:** Paraaortic lymph node metastasis

*Stage IV:* The carcinoma has extended beyond the true pelvis or has involved (biopsy proven) the mucosa of the bladder or rectum. A bullous edema, as such, does not permit a case to be allotted to stage IV.

- **IVA:** Spread of the growth to adjacent organs
- **IVB:** Spread to distant organs

## Appendix B

Main Matlab Code

1. **CNN**

```
% this demo will get MR data automatically and start a training on network

addpath('../../Training' , '../../mdCNN' , '../../utilCode' );

net = CreateNet('../../Configs/mnist.conf');

folder ='E:\After Defence\DatasetPerPatient\Binary\TrainBinary\';
dicomFiles = dir( fullfile(folder, '*.dcm' ));
y = length(dicomFiles);
I = zeros(288, 288, y, 'uint16');

for k = 1 : y
      % Read image
      filename = dicomread( fullfile( folder, dicomFiles(k).name ));
% I(:, :, k) =imresize(filename,[288, 288]);
     y=y+1;
end

folder1 = 'E:\After Defence\DatasetPerPatient\Both\Binary\Testing\';
dicomFiles1 = dir( fullfile(folder1, '*.dcm' ));
x = length(dicomFiles1);
I_test = zeros(288, 288, x, 'uint16');
for z = 1 : x
      % Read image.
```

```
      filename = dicomread( fullfile( folder1, dicomFiles1(z).name ));
%
I_test(:, :, z) = imresize(filename , [288, 288]);
  x=x+1;
end

%%%Binary

%%% Training 1960
labels = zeros( 1,1960 );   % 3 class
labels (1:800)=0;
labels (801:1360)=1;
labels (1361:1960)=2;

% labels = zeros( 1,1960 );   % 5 class
% labels (1:800)=0;
% labels (801:1040)=1;
% labels (1041:1360)=2;
% labels (1361: 1680)=3;
% labels (1681:1960)=4;

% labels = labels + 1;
%%%%%%%%%%%%%%%%%%%%

% labels_test = labels_test + 1;

%%%%%% Axial
% labels_test = zeros( 1,640);   %2 Class
% labels_test(1:320)=0;
% labels_test(321:640)=1;

% labels_test = zeros( 1,640);   %5 class
% labels_test(1:320)=0;
% labels_test(321:400)=1;
% labels_test(401:480)=2;
% labels_test(481:560)=3;
% labels_test(561: 640)=4;

labels_test = zeros( 1,640);   %3 class
labels_test(1:320)=0;
labels_test(321:480)=1;
labels_test(481:640)=2;

save bathymetry I labels I_test labels_test

outFile = fullfile(folder ,'bathymetry.mat');

save (outFile, 'I', 'labels', 'I_test', 'labels_test')
```

```
MNIST=load(outFile);

% start training

net   =  Train(MNIST,net,length(MNIST.I)*50); % 3,6,7,8,9,12,15,18 epocs

checkNetwork(net,Inf,MNIST,1);
```

**2. SVM**

```
%%%%%%%%%% Training Data
%%%%%%%%%% Training Data

cd 'E:\After Defence\DatasetPerPatient\Binary\TrainBinary'
folder ='E:\After Defence\DatasetPerPatient\Binary\TrainBinary\';

dicomFiles = dir( fullfile(folder, '*.dcm' ));
y = length(dicomFiles);

DF=[];

for i = 1:y

   J = dicomread(dicomFiles(i).name) ;  % Read image

   size_of_cropped_img = 200;
  [ny,nx] = size(J) ;
   C = round([nx ny]/2) ;
   x_cent = C(1);
   y_cent = C(2);
   centroide = [x_cent y_cent];

   xmin = x_cent-size_of_cropped_img/2;
   ymin = y_cent-size_of_cropped_img/2;

    I(:,:,1)= imcrop(J(:,:,1),[xmin ymin size_of_cropped_img size_of_cropped_img]);

[counts,x] = imhist(I);
stem(x,counts)
T = otsuthresh(counts);
BW = imbinarize(I,T);

stats =
regionprops(BW,'Area','Centroid','Perimeter','MajorAxisLength','MinorAxisLength','Eccentricity','Orientation
'); %%
Big = max( [stats.Area] );
Small =min([stats.Area] );
```

```
Axis1= stats.MajorAxisLength;
Axis2= stats.MinorAxisLength;
E=stats.Eccentricity;
O=stats.Orientation;

Centers = stats.Centroid;
Perimeter = stats.Perimeter;
M=mean(I ,'all');
S = std2(I );
GLCM= graycomatrix(I ,'Offset',[2 0]);
textures = getGLCMtextures(GLCM);

t1= textures.Energy;
t2= textures.Contrast;
t3= textures.Entropy;
t4= textures.Homogeneity;

FEAT= horzcat(1,[t1 t2 t3 t4 M S Big Small Centers Perimeter Axis1 Axis2 E O]);
DF= [DF;FEAT];

end

%
% labels = zeros( 1,1960 );   % 2 Class
% labels (1:800)=0;
% labels (801:1960)=1;
%

labels = zeros( 1,1960 );   % 3 class
labels (1:800)=0;
labels (801:1360)=1;
labels (1361:1960)=2;

% labels = zeros( 1,1960 );   % 5 class
% labels (1:800)=0;
% labels (801:1040)=1;
% labels (1041:1360)=2;
% labels (1361: 1680)=3;
% labels (1681:1960)=4;

labels = labels + 1;

%%%%%%%%%%%%%%%%%%%%%% MULTI CLASS SVM
%% Multi Class
N_train=1960;

TrainSet = DF(1:N_train,:);
GroupTrain = labels(1:N_train);
```

```
%%%%%%%%%%%%% Testing Data
%working
cd 'E:\After Defence\DatasetPerPatient\Both\Binary\Testing'
folder = 'E:\After Defence\DatasetPerPatient\Both\Binary\Testing\';
dicomFiles = dir( fullfile(folder, '*.dcm' ));
z = length(dicomFiles);

QF=[];

for i = 1:z

    B = dicomread(dicomFiles(i).name) ;  % Read image
    size_of_cropped_img = 200;
   [ny,nx] = size(B) ;
    C = round([nx ny]/2) ;
    x_cent = C(1);
    y_cent = C(2);
    centroide = [x_cent y_cent];
    xmin = x_cent-size_of_cropped_img/2;
    ymin = y_cent-size_of_cropped_img/2;

     K(:,:,1)= imcrop(B(:,:,1),[xmin ymin size_of_cropped_img size_of_cropped_img]);

[counts,x] = imhist(K);
stem(x,counts)
T = otsuthresh(counts);
BW = imbinarize(K,T);

stats =
regionprops(BW,'Area','Centroid','Perimeter','MajorAxisLength','MinorAxisLength','Eccentricity','Orientation
'); %,'Area','Centroid','Perimeter'
Big2 = max( [stats.Area] );
Small2 =min([stats.Area] );
Axis3= stats.MajorAxisLength;
Axis4= stats.MinorAxisLength;
E2=stats.Eccentricity;
O2=stats.Orientation;
Centers2 = stats.Centroid;
Perimeter2 = stats.Perimeter;

M2=mean(K ,'all');
S2 = std2(K );

GLCM= graycomatrix(K ,'Offset',[2 0]);
textures = getGLCMtextures(GLCM);
t1= textures.Energy;
```

```
t2= textures.Contrast;
t3= textures.Entropy;
t4= textures.Homogeneity;
FEAT= horzcat(1,[t1 t2 t3 t4 M2 S2 Big2 Small2 Centers2 Perimeter2 Axis3 Axis4 E2 O2]);
QF= [QF;FEAT];

end

TestingSet=QF;
% %
% labels_test = zeros( 1,640);   %2 Class
% labels_test(1:320)=0;
% labels_test(321:640)=1;
% labels_test = labels_test + 1;

labels_test = zeros( 1,640);   %3 class
labels_test(1:320)=0;
labels_test(321:480)=1;
labels_test(481:640)=2;
labels_test = labels_test + 1;

% %
% labels_test = zeros( 1,640);   %5 Class
% labels_test(1:320)=0;
% labels_test(321:400)=1;
% labels_test(401:480)=2;
% labels_test(481:560)=3;
% labels_test(561: 640)=4;
% labels_test = labels_test + 1;
%

N_Test=640;
GrouppTest = labels_test(1:N_Test);
 % %  With MULTISM FUNCTION
[Output,Scor]= multisvm(TrainSet,GroupTrain,TestingSet);
Predict= transpose(Output);
g2=Output;
g1=transpose(labels_test);
[GC,GR] = groupcounts(Predict);

 Accuracy = sum( GrouppTest==Predict) / numel(labels_test)*100;  %%working
figure
C = confusionchart(GrouppTest,Predict, ...
    'ColumnSummary','column-normalized', ...
    'RowSummary','row-normalized');
C.ColumnSummary = 'column-normalized';
C.RowSummary = 'row-normalized';
C.Title = 'SVM Confusion Matrix';
```

### 3. KNN

```
% %%%%%%%%%% Training Data

cd 'E:\After Defence\DatasetPerPatient\Binary\TrainBinary'
folder ='E:\After Defence\DatasetPerPatient\Binary\TrainBinary\';

dicomFiles = dir( fullfile(folder, '*.dcm' ));
y = length(dicomFiles);

DF=[];

for i = 1:y

    J = dicomread(dicomFiles(i).name) ;  % Read image
    size_of_cropped_img = 200;
   [ny,nx] = size(J) ;
    C = round([nx ny]/2) ;
    x_cent = C(1);
    y_cent = C(2);
    centroide = [x_cent y_cent];
    xmin = x_cent-size_of_cropped_img/2;
    ymin = y_cent-size_of_cropped_img/2;

     I(:,:,1)= imcrop(J(:,:,1),[xmin ymin size_of_cropped_img size_of_cropped_img]);

[counts,x] = imhist(I);
stem(x,counts)
T = otsuthresh(counts);
BW = imbinarize(I,T);

stats =
regionprops(BW,'Area','Centroid','Perimeter','MajorAxisLength','MinorAxisLength','Eccentricity','Orientation
'); %%
Big = max( [stats.Area] );
Small =min([stats.Area] );
Axis1= stats.MajorAxisLength;
Axis2= stats.MinorAxisLength;
E=stats.Eccentricity;
O=stats.Orientation;

Centers = stats.Centroid;
Perimeter = stats.Perimeter;
M=mean(I ,'all');
S = std2(I);
GLCM= graycomatrix(I,'Offset',[2 0]);
textures = getGLCMtextures(GLCM);
```

```
t1= textures.Energy;
t2= textures.Contrast;
t3= textures.Entropy;
t4= textures.Homogeneity;

FEAT= horzcat(1,[t1 t2 t3 t4 M S Big Small Centers Perimeter Axis1 Axis2 E O]);
DF= [DF;FEAT];

end
%
labels = zeros( 1,1960 );   % 2 Class
labels (1:800)=0;
labels (801:1960)=1;

% labels = zeros( 1,1960 );   % 3 class
% labels (1:800)=0;
% labels (801:1360)=1;
% labels (1361:1960)=2;

%
% labels = zeros( 1,1960 );   % 5 Class
% labels (1:800)=0;
% labels (801:1040)=1;
% labels (1041:1360)=2;
% labels (1361: 1680)=3;
% labels (1681:1960)=4;

labels = labels + 1;

%%%%%%%%%%%%%%%%%%%%%% MULTI CLASS SVM
N_train=1960;
TrainSet = DF(1:N_train,:);
GroupTrain = labels(1:N_train);

 %%%%%%%%%%%%% Testing Data
%
cd 'E:\After Defence\DatasetPerPatient\Both\Binary\Testing'
folder = 'E:\After Defence\DatasetPerPatient\Both\Binary\Testing\';

dicomFiles = dir( fullfile(folder, '*.dcm' ));
z = length(dicomFiles);

QF=[];

for i = 1:z

   B = dicomread(dicomFiles(i).name) ;  % Read image
size_of_cropped_img = 200;
```

```
   [ny,nx] = size(B) ;
    C = round([nx ny]/2) ;
    x_cent = C(1);
    y_cent = C(2);
    centroide = [x_cent y_cent];

    xmin = x_cent-size_of_cropped_img/2;
    ymin = y_cent-size_of_cropped_img/2;

     K(:,:,1)= imcrop(B(:,:,1),[xmin ymin size_of_cropped_img size_of_cropped_img]);
[counts,x] = imhist(B);
stem(x,counts)
T = otsuthresh(counts);
BW = imbinarize(K,T);

stats =
regionprops(BW,'Area','Centroid','Perimeter','MajorAxisLength','MinorAxisLength','Eccentricity','Orientation
'); %,'Area','Centroid','Perimeter'
Big2 = max( [stats.Area] );
Small2 =min([stats.Area] );
Axis3= stats.MajorAxisLength;
Axis4= stats.MinorAxisLength;
E2=stats.Eccentricity;
O2=stats.Orientation;
Centers2 = stats.Centroid;
Perimeter2 = stats.Perimeter;

M2=mean(K ,'all');
S2 = std2(K );

GLCM= graycomatrix(K ,'Offset',[2 0]);
textures = getGLCMtextures(GLCM);
t1= textures.Energy;
t2= textures.Contrast;
t3= textures.Entropy;
t4= textures.Homogeneity;
FEAT= horzcat(1,[t1 t2 t3 t4 M2 S2 Big2 Small2 Centers2 Perimeter2 Axis3 Axis4 E2 O2]);
QF= [QF;FEAT];

end

TestingSet=QF;
% %
labels_test = zeros( 1,640);   %T2 New Testing
labels_test(1:320)=0;
labels_test(321:640)=1;
labels_test = labels_test + 1;
%
```

```
%3 class
% labels_test = zeros( 1,640); %%%%%%%% 3 Class
% labels_test(1:320)=0;
% labels_test(321:480)=1;
% labels_test(481:640)=2;
% labels_test = labels_test + 1;


% labels_test = zeros( 1,640);   %5 Class
% labels_test(1:320)=0;
% labels_test(321:400)=1;
% labels_test(401:480)=2;
% labels_test(481:560)=3;
% labels_test(561: 640)=4;
% labels_test = labels_test + 1;

N_Test= 640;
GrouppTest = labels_test(1:N_Test);
%%%%%%%%% KNN

 KNNModel = fitcknn(TrainSet,GroupTrain,'NumNeighbors',25,'Standardize',1);
%  predictedY = resubPredict(KNNModel);
[label,score,cost] = predict(KNNModel,TestingSet);

figure
C = confusionchart(GrouppTest,label, ...
    'ColumnSummary','column-normalized', ...
    'RowSummary','row-normalized');
C.ColumnSummary = 'column-normalized';
C.RowSummary = 'row-normalized';
C.Title = 'KNN Confusion Matrix';

g2=label;
g1=transpose(labels_test);

Accuracy = sum( g1==g2) / numel(labels_test)*100;
CVKNNMdl = crossval(KNNModel);
loss = kfoldLoss(CVKNNMdl);
% fprintf('\n The Classification Accuracy is %0.4f', Accuracy);
%
CrossValAccuracy= (100 - (loss*100));

fprintf('\n The Classification Accuracy with Cross Validation is %0.4f', CrossValAccuracy);
fprintf('\n The Classification Accuracy is %0.4f', Accuracy);
```

# Appendix C

Ref. No. 8 M23/435

Date: 10/2/2020

**Institutional Review Board (IRB) of St. Paul's Hospital Millennium Medical College (SPHHMC)**

**Ethical Clearance**

**Research Title:** - Staging of Cervical Cancer Using Convolution Neural Network for Treatment and Prognosis Decision Support System

**Principal Investigator:** - Kidist Kebede

The IRB of SPHMMC has reviewed the above mentioned research proposal and made the following decision:

- Approved:- ✓
- Approved with recommendation:- _____
- Approved on condition :- _____
- Disapproved:- _____

The decision is valid for 12 months and the research should be conducted in compliance with the protocol/proposal approved by the IRB of SPHMMC. Any subsequent revision/amendment of the protocol/proposal needs approval before conduct of the research. The researcher should also submit written summaries of the research status to the IRB every 03 months. Upon the conclusion of the study, manuscripts and thesis work to the final/completed research project needs to be submitted to the IRB.

IRB Chair:

Signature: _____

Mahlteme Bekele (MD)
Associate Professor Research
Directorate Director

Date: February 18, 2020

**Cc:**

- Vice Provost for Academic and Research
- IRB
- Kidist Kebede
  SPHMMC