Ovarian Cancer detection using Image Processing and Machine Learning

Chaitra B Kamat[1], Vandana A S[2], Bindu S S[3], Ranjana N[4]
Guide: Bhuvaneshwari K V, Assistant Professor
Bapuji Institute of Engineering and Technology
(E-mail: chaitrabkamat@gmail.com
vandanask97@gmail.com
bindussfeb22@gmail.com
ranjana.n1997@gmail.com)

Abstract— Nowadays cancer is the second leading cause of death all over the world. Cancer is one such disease which can be categorized into various subtypes. Ovarian cancer is a type of cancer that begins in the ovaries and goes undetected until it spreads within the pelvis and abdomen. Early diagnosis and treatment are necessary for improving the patients cure rate and prolonging their survival. To overcome these problems image processing and machine learning techniques are widely used in medical areas for improving earlier detection and treatment stages, in which the time factor is very important to discover the disease in the patient as fast as possible, Thus, these techniques provide accurate classification of benign and malignant tumors and prevent patients from undergoing unnecessary treatments.

Keywords*: Image Processing, Machine Learning*

I.INTRODUCTION

Cancer develops when abnormal cells in a part of the body (in this case, the ovary) begin to grow uncontrollably. This abnormal cell growth is common among all cancer types. In women ages 35-74, ovarian cancer is the fifth leading cause of cancer-related deaths. An estimated one woman in 75 will develop ovarian cancer during her lifetime. The American Cancer Society estimates that there will be over 22,280 new cases of ovarian cancer diagnosed this year and that more than 14,240 women will die from ovarian cancer this year. Cancer is the second leading cause of death, alongside heart disease, in both developed and developing countries. Cancer has been characterized as a heterogeneous disease consisting of several different subtypes. Early diagnosis and prognosis of a sort of disease have turned into a need in malignancy explore, as this can encourage the resulting clinical administration of patients.

At present, there are top notch open databases including both molecular estimations of tumor, and additionally clinical information on patients. By applying machine learning methods to these databases, it is possible even for non-experimenters to generate plausible assumptions that are supported by the data, which can then be validated on one or more independent data sets. As a result, these techniques have been used to model the progression and treatment of cancerous diseases. Also, the ability of machine learning tools to detect key features from complex data sets reveals their importance. Given the growing trend in the application of machine learning methods in cancer research, in this paper, the most recent publications that use these techniques as an objective to model cancer risk or outcomes are reviewed, and their applications to ovarian cancer are discussed.

When one is diagnosed and treated in the earliest stages, the five-year survival rate is over 90 percent. Due to ovarian cancer's non-specific symptoms and lack of early detection tests, about 20 percent of all cases are found early, meaning in stage I or II. If caught in stage III or higher, the survival rate can be as low as 28 percent. Due to the nature of the disease, each woman diagnosed with ovarian cancer has a different profile and it is impossible to provide a general prognosis. Ovarian cancer is a disease in which, depending on the type and stage of the disease, malignant (cancerous) cells are found inside, near, or on the outer layer of the ovaries. An ovary is one of two small, almond-shaped organs located on each side of the uterus that store eggs, or germ cells, and produce female hormones estrogen and progesterone.

Most women with ovarian cancer are diagnosed with advanced-stage disease (Stage III or IV). This is because the symptoms of ovarian cancer, particularly in the early stages, often are not acute or intense, and present vaguely. In most cases, ovarian cancer is not detected during routine pelvic exams, unless the doctor notes that the ovary is enlarged. The sooner ovarian cancer is found and treated, the better a woman's chance for survival. It is important to know that early stage symptoms can be difficult to detect, though are not always silent. As a result, it is important that women listen to their bodies and watch for early symptoms that may present.

Thus, Image Processing and Machine learning techniques can be used in the detection of ovarian cancer. In recent years, the image processing mechanisms are widely used in several medical areas for improving earlier detection and treatment stages, in which the time factors is very important to discover the disease in the patient as fast as possible, especially in various cancers such as ovarian cancer. Early detection is very important for successful treatment. Image Processing methods such as Segmentation play a significant role in detection of cancer.

Machine Learning, subfield of Artificial Intelligence has been widely used in the development of predictive models in order to support effective decision making. In cancer research, these techniques could be used to identify different patterns in a dataset and consequently predict whether a cancer is malignant or benign. There are two primary basic sorts of machine learning strategies known as supervised learning and unsupervised learning. In supervised learning, a named set of preparing information is utilized to gauge or guide the information to the coveted yield. On the other hand, under the unsupervised learning strategies, no named illustrations are given, and there is no thought of the yield amid the learning procedure. This procedure can be considered as a classification problem in supervised learning.

II.EXISTING SYSTEM

Between 70 and 75% of ovarian carcinomas are not discovered until they have reached an advanced stage III or later. Efforts should therefore be concentrated on earlier diagnosis. Ovarian cancer is not an entirely silent disease. Today, it is known that there are key symptoms which, depending on their frequency and intensity, can serve as warning signs to clinicians and patients. Mass screening for ovarian cancer is not currently possible because of a lack of specific markers for use in biological and imaging techniques, although new markers are now being developed.

Screening every six or twelve months with the CA 125 blood test plus a transvaginal ultrasound is restricted to women at risk. Certain teams have proposed preventive bilateral adnexectomy for such women. The ovary is a complex organ subjected to a hormonal environment and affected by immune system dysfunctions.

II. PROPOSED METHOD

The training and test data must follow the same probability distribution is one of the essential premises of machine learning. The purpose behind this suspicion is plainly obvious: if an algorithm is formed on a data set and tested on an entirely different dataset, it generally cannot have good performance on the test data. In the early years, levels of gene expression were measured using microarrays. This approach has resulted in relatively inaccurate and fairly no repeatable estimations. The present pattern is to supplant microarrays by something many refer to as RNA-seq. Not at all like microarray estimations, are RNA-seq estimations to a great degree precise. Along these lines it is important to fit both the legacy microarray information with more up to date RNA-seq information. So, with new RNA-seq data, it is essential to harmonize both the existing microarray data. There is a generally new region in machine learning known as "transfer learning" in which the preparation information and test information are permitted to have different probability distributions.

IV.METHODOLOGY

Initially we input the dicom images and carry out the pre-processing steps such as segmentation so as to find which region of the organ has been affected. Then we extract some of the features of that affected region. Later based on machine learning techniques we predict as in which stage the cancer is in and also determine the type of cancer(malignant or benign)

V. LITERATURE SURVEY

1. Suthamerthi Elavarasu, Viji Vinod, Elavarasan Elangovan:"Machine Learning Applications in Ovarian Cancer Prediction", 2017:

By applying machine learning methods to these databases, it is possible even for non-experimenters to generate plausible assumptions that are supported by the data, which can then be validated on one or more independent data sets. As a result, these techniques have been used to model the progression and treatment of cancerous diseases. Given the growing trend in the application of machine learning methods in cancer research, in this paper, the most recent publications that use these techniques as an objective to model cancer risk or outcomes are reviewed, and their applications of ovarian cancer. [1]

2. Sharmistha Bhattacharjee, Yumnam Jayanta Singh, Dipankar Ray:"Comprative performance analysis of machine learning classifiers on ovarian cancer dataset", 2017:

Machine learning classifiers are applied on the Ovarian Cancer Dataset. Using Machine learning techniques, data is classified in different categories to identify benign and malignant cancerous cells and a comparative study has been done to identify the most suitable technique under different operational conditions and dataset. [2]

3. G. Vasavi, S. Jyothi: "Classification and detection of ovarian cysts in ultrasound images", 2017:

Ovarian cancer must be diagnosed properly to analyse the type of tumors and malignant cysts in the body.Different follicles and cysts are classified in this paper for better understand of the disease Some of Digital Image processing methods are analysed to detect the ovarian cysts from ultrasound images of ovaries. Benign images are considered to analyse the extra masses. Some of Digital Image processing methods are analysed to detect the ovarian cysts from ultrasound images of ovaries. [3]

4. T. M. Shahriar Sazzad, L. J. Armstrong, A. K. Tripathy:"An Automated Detection Process to Detect Ovarian Tissues Using Type P63 Digitized Color Images", 2015:

Automated detection of ovarian reproductive tissues is one of the important diagnosis interests for pathologists these days. One of the popular diagnosis preferences to identify ovarian tissues is ultrasound scanner. However, due to different shape, size and colour, identification of ovarian tissues is a challenging task for ultrasound scanners as it process gray scale images. [4]

5. Priya Darshini Velusamy, Porkumaran Karandharaj: "Medical Image Processing schemes for cancer detection", 2014:

In this paper the survey has been proposed on "Medical Image processing schemes for cancer detection". This paper provides details of various techniques that reveal how hybrid intelligent approaches are applied to different categories of cancer detections and treatment. By reviewing all the methods available in the literature, it is found that each method has its own merits and demerits. It provides overview of various states of methods available in the cancer detection and comparison analysis of each method. [5]

6. Kiruthika, V., Ramya, M.: Automatic segmentation of ovarian follicle using K-means clustering. In: 2014 Fifth International Conference on Signal and Image Processing (ICSIP), pp. 137–141 (2014):

Automatic detection of human ovarian follicles has been of increasing interest in recent years and is a significant area of women's health. Improper development of ovarian follicles has been an important reason for infertility in women. Currently, detection of ovarian follicle is done through diagnostic imaging technique called ultrasonography. Follicles differ in shape and color. Further, the camouflaging characteristic of ultrasound images and the presence of speckle noise make the follicle detection a challenging task. In this paper, a novel method for automatic recognition of follicles in ultrasound images is proposed. Discrete wavelet transform based k-means clustering is proposed.

VI. SYSTEM REQUIREMENTS AND SPECIFICATION

Functional Requirements

1. Image dataset: A machine learning image dataset is used for the prediction. The image recognition dataset for developing unsupervised feature learning.

2. Extracting tumor region: Extraction and description of tumor region from the ovarian MRI image using segmentation techniques. Image segmentations refer to partitioning of images into multiple regions of images through which information that are extracted.

3. Finding boundary of the tumor: Ovarian tumor boundary is one of the challenging task in the medical image processing.

4. GUI: The graphical user interface is a form of user interface that allows users to interact with electronic devices through graphical icons and visual indicators.

Non -functional requirements

1. Availability: The software for extraction of ovarian tumor from CT scan images can be available in all the system who have MATLAB installed.

2. Reliability: The software attempts to ensure appropriate content but assume no responsibility for external manipulation.

3. Accuracy: Effect of using the same versus different order for second reading of screening mammograms on rates of ovarian cancer.

4. Performance: Patient performance status is an important part of cancer care and treatment. Image processing and Machine learning techniques provide better performance in diagnosis.

Hardware requirements

1. RAM - 4GB
2. Hard Disk - 10GB

Software requirements

1. Operating system - Windows 7 or higher version
2. Platform - Anaconda Navigator, MySQL, Matlab
3. Programming Languages - Python/R language

ACKNOWLEDGMENT

We are grateful to many people. First, We whole heartedly express our gratitude to our respected Principal Dr Nataraja, Director Prof Y. Vrushabhendrappa, HOD Dr Poornima B and Asst. Prof Ms. Bhuvaneshwari K V for encouragement, valuable guidance and constant support. Finally, we thank GSSSIET, Mysuru for providing opportunity.

REFERENCES

1. Suthamerthi Elavarasu, Viji Vinod, Elavarasan Elangovan:"Machine Learning Applications in Ovarian Cancer Prediction", 2017:
2. Sharmistha Bhattacharjee, Yumnam Jayanta Singh, Dipankar Ray:"Comprative performance analysis of machine learning classifiers on ovarian cancer dataset", 2017:
3. G. Vasavi, S. Jyothi: "Classification and detection of ovarian cysts in ultrasound images", 2017:
4. T. M. Shahriar Sazzad, L. J. Armstrong, A. K. Tripathy:"An Automated Detection Process to Detect Ovarian Tissues Using Type P63 Digitized Color Images", 2015:
5. Priya Darshini Velusamy, Porkumaran Karandharaj: "Medical Image Processing schemes for cancer detection", 2014:

AUTHOR PROFILE

1. CHAITRA BHAIRAV KAMAT

- 1st place in Business Model Presentation in the State Level Hackathon-2018, Shivamogga.
- 2nd place in Papyrus- A State Level Technical Paper Presentation on "Touch Your Sound Technology".

2. VANDANA A S

3. BINDU S S

- Participated in 'Smart India Hackathon-2018'

4. RANJANA N