

Harnessing the Power of Image Analysis and Machine Learning for Accurate Oncology Prognosis

Deeksha Shetty, Bhoomika L, Srushti M, Bhoomika D L,
Amit Kumar K

Dept. of Information Science and Engineering PESITM, Shivamogga
Dept. of Information Science and Engineering PESITM, Shivamogga
Dept. of Information Science and Engineering PESITM, Shivamogga
Dept. of Information Science and Engineering PESITM, Shivamogga
Dept. of Information Science and Engineering PESITM, Shivamogga

Date of Submission: 01-04-2023

Date of Acceptance: 10-04-2023

ABSTRACT-Cancer has been described as a diverse illness with a wide range of subgroups. Since it can simplify later clinical patient treatment, early cancer detection and prognosis have become essential in cancer research.

Keywords: Image Processing, Oncology, Machine Learning, Segmentation, Classifier.

I. INTRODUCTION

Cancer research has seen a constant change throughout the past few decades [1]. Researchers have used a variety of techniques, such as early-stage screening, to identify cancer types before symptoms appear. Also, they have created fresh methods for anticipating the results of cancer treatments in advance. Large volumes of cancer data have been gathered and are available to the medical research community thanks to the development of new technologies in the field of medicine. Nonetheless, one of the most fascinating and difficult problems for doctors is making an accurate prognosis of a disease's course. Hence, ML techniques have gained popularity as a tool for medical researchers. These methods are capable of extracting patterns and correlations from large datasets, as well as efficiently forecasting the course of a particular cancer type.

We give a review of research that employ these approaches in relation to the prediction and prognosis of cancer in light of the relevance of personalized medicine and the emerging trend on the application of ML techniques. In these research, prognostic and predictive factors that may be independent of a particular treatment or that are combined to direct therapy for cancer patients,

respectively, are taken into consideration [2]. Also, we talk about the many ML techniques being utilised, the different kinds of data they incorporate, and the effectiveness of each suggested method, as well as their advantages and disadvantages.

The incorporation of mixed data, including clinical and genetic data, is a clear trend in the suggested works. Nevertheless, a recurring issue that we identified in several works is the absence of external validation or testing about the accuracy of their models. There is no doubt that using ML techniques might increase the precision of cancer susceptibility, recurrence, and survival prediction. According to [3], the use of ML approaches has greatly increased cancer outcome prediction accuracy over the last years by 15%–20%.

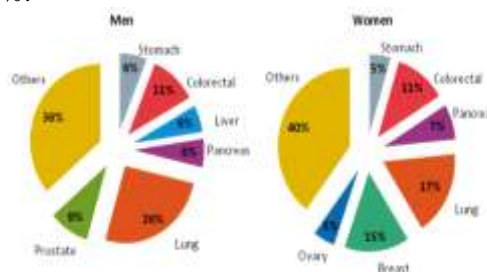


Fig 1. Global Cancer Statistics

An estimated 20 million new instances of cancer and 10 million cancer-related deaths occur each year globally. During the next 20 years, the burden of cancer is expected to rise by around 60%, putting extra strain on health systems, individuals, and communities. The projected global burden will

rise to about 30 million new cancer cases by 2040, with the largest increase in low- and middle-income countries.

Several studies have been published in the literature and are based on various approaches that may enable the early diagnosis and prognosis of cancer [4], [5], [6], and [7]. These publications address methods for analysing circulating miRNAs, which have been identified as a promising class for cancer diagnosis and characterization. Nevertheless, these techniques have limited sensitivity when used for early-stage screening and have trouble differentiating between benign and malignant tumours. In [8] and [9], many elements of the prediction of cancer prognosis based on gene expression profiles are covered. These papers outline the advantages and disadvantages of microarrays for predicting the course of cancer. Despite the fact that gene signatures have the potential to greatly improve cancer patient prognosis, little progress has been achieved in using them in clinical settings. Nevertheless, investigations with bigger data sets and more sufficient validation are required before gene expression profiling may be employed in clinical practice.

In the United States, cancer-related fatalities are predicted to total 609,820 in 2023, or 1,670 deaths every day. For both men and women, lung, breast, and colon cancers account for the majority of fatalities.

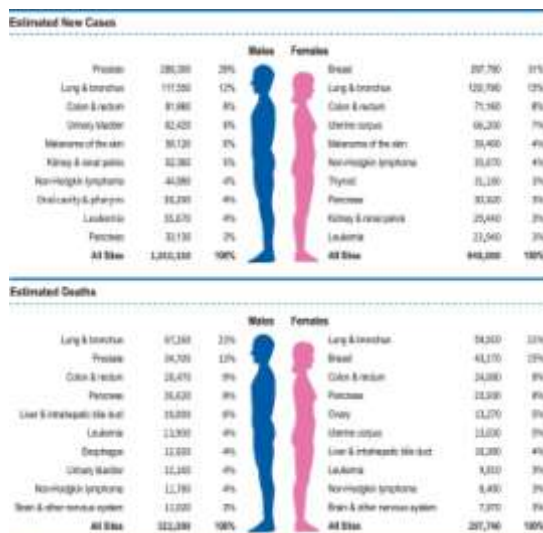


Fig 2. The different cancer types and new cases estimation

In the current study, only studies using ML techniques to model cancer diagnosis and prognosis are presented. Computer methods known as machine learning (ML) enable computers to

recognize patterns in data and predict future outcomes using precedents. Machine learning (ML), a term initially used by Arthur Samuel, is regarded as a significant area of artificial intelligence and was introduced by John McCarthy. It is described as the ability of computers to carry out tasks like those of human intellect. The typical purpose of ML is to discover analytical patterns in data and generate generalizations (predictions) based on samples that have already been analyzed. Hence, it is a topic that focuses on data mining and performance prediction and is closely tied to cognitive psychology, neurology, computational, and statistical concepts.

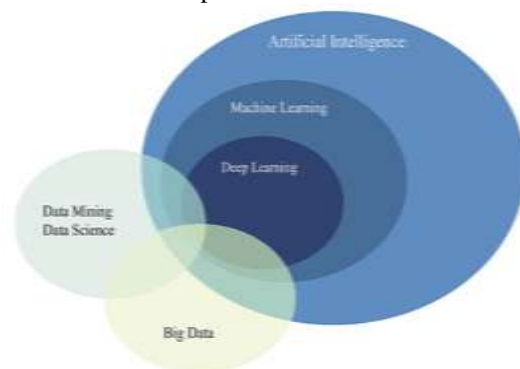


Fig 3. Relation between ML, DL, DS

Keep in mind that machine learning is the computational area of AI designed to enable computers to do jobs outside of the scope of their original programming big data analytics and data mining analytics.

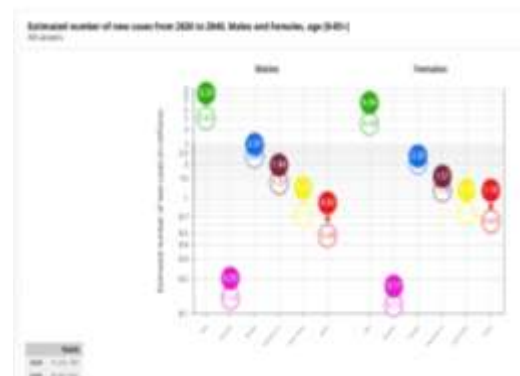


Fig 4. Estimated number of new cases from 2020-2040

II. LITERATURE SURVEY

Early research in this field focused on categorizing histological pictures and detecting diseases using computer-assisted disease diagnosis (CAD). To acquire precise data, a model for automated cancer diagnosis is utilized using a

number of parameters, including the region of interest, variance of information (VOI), and false error rate. The sample picture is a magnetic resonance imaging (MRT) image, and identification is carried out using K-Nearest Neighbor (KNN) and Linear Discriminate Analysis (LDA) (LDA). Support machine vectors are also used to sample for thermographs and mammography (SVM).

We think that the application of machine learning approaches for cancer prediction and prognosis has resulted in an increase in the number of published research given the rising relevance of predictive medicine and the growing dependence on machine learning to make predictions. A thorough analysis seemed fascinating, so we decided to do one. The objective is to identify the various machine learning techniques used, the training data types that were incorporated, the endpoint predictions that were made, the various cancer types that were investigated, and the effectiveness of these techniques in predicting cancer susceptibility or patients. To recognize significant trends in the overall performance of the outcome. Surprisingly, we discovered that the majority of research focused on three "predictive" foci or clinical endpoints when discussing cancer prognosis and prediction, Cancer survival prediction, Cancer recurrence prediction.

Moreover, ML offers prescriptive insights. Treatment choices and patient monitoring can be influenced by individualized forecasts of therapeutic response to alternative medicines and their possible negative effects. A significant part of this has been played by genomic data. The capacity to predict drug susceptibility on a wide scale using genetic profiles has been made possible by the increased availability of cell line data. In more focused interactions, such as leucovorin, fluorouracil, and oxaliplatin responses in patients with colorectal cancer, genomic information has also been utilized to predict clinical response metrics. Patients getting neoadjuvant chemotherapy have had their treatment outcomes predicted using machine learning (ML). Breast cancer has been treated with a mix of clinical and imaging techniques, while non-small cell lung cancer (NSCLC) has been treated with radiomics. Several investigations have been carried out to find negative consequences of therapy, either at the medication level or patient level.

ML can be used to evaluate the response of the tumor. Traditionally, RECIST's two-dimensional tumor measures have been used for this. The requirement to employ characteristics that radiologists can measure leads to the dependence

on 2D measures. This strategy has limitations, and experts point out that RECIST may not precisely measure progress in outcomes. ML has been used to automatically identify RECIST criteria in patients with NSCLC since it has shown promise in diagnostic imaging. Serial CT scans for NSCLC and volumetric measures of brain tumours utilising magnetic resonance imaging have been proposed as alternatives to his RECIST for evaluating response.

In summary, ML has potential in cancer. It may be used to identify risk groups, forecast prognoses, guide treatment options, and support early detection and diagnosis. We can better understand cancer and how it affects people by using data-driven techniques because there is a wealth of patient data available. While ML has several organizational and technological difficulties, it is a commendable undertaking to revolutionize cancer therapy.

III. PROPOSED SYSTEM

Cancer detection is a multistep process. Patients often go to the doctor because of some symptoms. Cancer might be identified by screening or by coincidence. Getting a selection of microscopic biopsy pictures is the first step in the first phase. Each image goes through a number of preparation steps, such as noise reduction and enhancement.

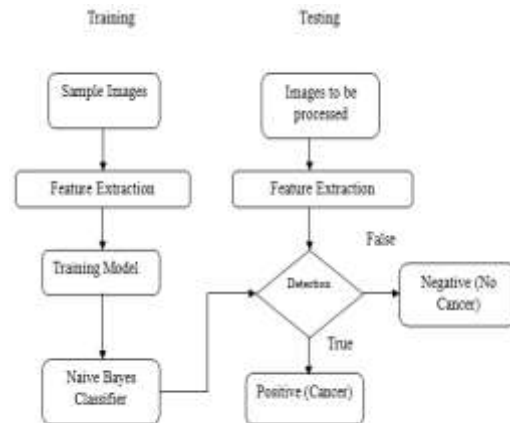


Fig 5. Flowchart of cancer prediction

The above mentioned diagram indicates the procedure for cancer prediction:

- Training data and test data are separated from the dataset. Also, he has two stages: a testing phase and a training phase.
- During the training phase, it is determined how data and characteristics relate to one another. For systems that extract certain traits, example data is given. Based on these extracted attributes, a model is produced. To train the model, a classifier is

utilized to categorize all provided data.

- The same characteristics that were discovered in the training phase are retrieved from photos in the test phase.
- Color, shape, and positioning are compared and used to sort new photos. At this phase, the picture is identified and shown as either favorable or unfavorable.
- Results that are either positive or negative depend on whether the cells are malignant.

IV. METHODOLOGY

The block diagram contains various steps:

- After the biopsy, microscopically tested images are taken as input. The image is enhanced and denoised before segmentation.
- Segmentation is done based on input images containing nuclear, cytoplasmic, and other features. They are segmented by regions, thresholds or clusters and a special algorithm is applied.

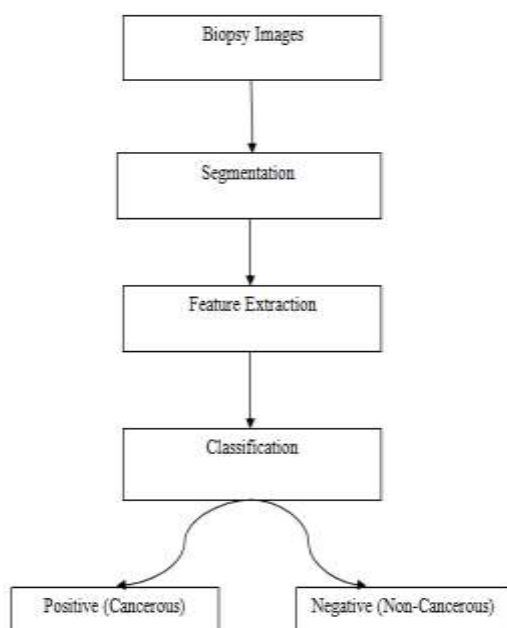


Fig6. Architecture diagram of cancer detection

- From segmented pictures, characteristics are extracted that are based on form and morphology and may be interpreted physiologically and therapeutically. This contains color-based features, grayscale texture characteristics, and features that are dependent on color.

ML for Image Processing in Oncology

Throughout the past few decades, clinical usage of medical imaging has grown significantly. Recent advancements in the ML community and enormous progress in quantitative imaging techniques have demonstrated even greater

potential and application.

Functional (molecular) imaging, such as positron emission tomography (PET), single photon emission computed tomography (SPECT), and diffusion-weighted MRI, are the two primary categories of image acquisition. SPECT/CT, PET/CT, and PET/MRI are examples of multimodal imaging techniques that have been developed to combine the benefits of anatomical resolution and functional tissue information. Images were encoded with patient-specific information about the tumor, tumor environment, and genotype using all of the tools that were accessible. This information may be assessed for diagnosis, prognosis, and prediction of oncology result. We can offer helpful information.

This makes it possible to ask: "How can we find underlying biological links in these massive volumes of imaging data?" It extracts more information, albeit in a rather qualitative way, on top of the already difficult process that physicians employ to examine and interpret medical pictures for everyday judgements (such as tumour borders and its heterogeneity, etc.).

Thankfully, high-throughput quantitative feature extraction has made it possible to convert digital medical pictures into useful high-dimensional data. This is made possible by the development of powerful pattern recognition algorithms and statistical learning tools. This has shown that cancer precision medicine has a lot of potential. The field of radiomics is the result of the following study of the transformation of medical pictures into a large number of complex characteristics and linking these features to biological endpoints and clinical outcomes.

V. EXPECTED RESULT

In this paper, we addressed a technique for automatically identifying and classifying cancer from microscopic biopsy pictures. Images of microscopic biopsies are imported from programme files. The image is read via image processing, and a CNN technique is used to segment it. Images are trained and tested using machine learning. Test the image and describe the outcome as either favorable or unfavorable. This process is quicker and accurately anticipates outcomes.

REFERENCES

- [1]. M. P. Chander, M. V. Rao, and T. Rajinikanth, "Detection of lung cancer using digital image processing techniques: a comparative study," International Journal of Medical Imaging,

- vol. 5, no. 4, pp. 59-62, 2017. doi: 10.11648/j.ijmi.20170505.12.
- [2]. S. S. Priya and B. Ramamurthy, "Lung cancer detection using image processing techniques," *Research Journal of Pharmacy Technology* vol. 11, no. 5, pp. 2045-2049, 2018. <http://dx.doi.org/10.5958/0974-360X.2018.00379.7>.
- [3]. Ram MurtiRawat ,Shivam Panchal ,Vivek Kumar Singh ,Yash Panchal "Breast Cancer Detection Using K-Nearest Neighbors, Logistic Regression and Ensemble Learning" (2020).
- [4]. Ghoshal S, Rigney G, Cheng D, et al. Institutional surgical response and associated volume trends throughout the COVID-19 pandemic and postvaccination recovery period. *JAMA Netw Open*. 2022;5(8):e2227443. doi:10.1001/jamanetworkopen.2022.27443
- [5]. Yabroff KR, Wu XC, Negoita S, et al. Association of the COVID-19 pandemic with patterns of statewide cancer services. *J Natl Cancer Inst*. 2022;114(6):907-909.
- [6]. Chen R, Aschmann HE, Chen YH, et al. Racial and ethnic disparities in estimated excess mortality from external causes in the US, March to December 2020. *JAMA Intern Med*. 2022;182(7):776-778. doi:10.1001/jamainternmed.2022.1461
- [7]. Woolf SH, Chapman DA, Sabo RT, Zimmerman EB. Excess deaths from COVID-19 and other causes in the US, March 1, 2020, to January 2, 2021. *JAMA*. 2021;325(17):1786. doi:10.1001/jama.2021.5199
- [8]. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: Incidence-SEER Research Data with Delay-Adjustment, 8 Registries, Malignant Only (1975-2019), based on the November 2021 submission. National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Surveillance Systems Branch; 2022.