A ADVANCED JOURNAL IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING TRENDS IN BIOLOGICAL SCIENCES
Open Journal

Review

Machine Learning and Deep Learning for Cancer Diagnosis via Logistic Regression: The Significance of Artificial Intelligence

Subash Kumar*

Bachelor of Engineering specialized in Computer Science, Anna University, India

Corresponding author
Subash Kumar
Bachelor of Engineering specialized in Computer Science, Anna University, India

Article information
Received: March 21st, 2022; **Revised:** July 31st, 2022; **Accepted:** August 20th, 2022; **Published:** October 18th, 2022

Cite this article
Kumar S. Machine learning and deep learning for cancer diagnosis via logistic regression: The significance of artificial intelligence. 2022; 1(1).
doi: https://doi.org/10.70705/ppp.bioai.2022.v01.i01.pp14-17

ABSTRACT

Artificial intelligence (AI) may be defined in a variety of ways. "The Study of the computations that make it possible to perceive, reason, act and predict the future possible outcomes" is one definition of artificial intelligence ("AI"). Artificial intelligence (AI), often known as machine intelligence, is a subfield of computer science that focuses on creating intelligent computer programs that can do tasks normally performed by people.We need high-quality computer assistance to make better use of our material and human resources as the world becomes more complicated. Some examples of potential applications include the following: the use of computer-controlled robots to manage pests, trim trees, and harvest mixed crops selectively in the agricultural sector. The tedious and sometimes hazardous tasks of assembly, inspection, and maintenance should be handled by computer-controlled robots in the manufacturing industry.Medical professionals should be able to use computers to aid in diagnosis, patient status monitoring, therapy management, and even bed making.In the realm of household tasks, computers need to provide gastronomic and grocery store recommendations, sweep, mow, wash, and tend to any necessary repairs.In the classroom, computers shouldn't only respond to students' mistakes; they should comprehend the reasons behind them. Students can better grasp physics and music with the use of computers that depict planetary orbits and play musical scores ([1]). Computers would be the greatest gift to humanity if they could aid medical professionals with diagnosis, track patients' vitals, and provide guidance to interdisciplinary teams.

Keywords
Machine learning; Deep learning; Artificial intelligence; Medicine; Lung cancer; Breast cancer; Cancer prognosis and prediction; Data science; Logistic regression; Regression; Supervised machine learning.

INTRODUCTION

The most recent figures on the worldwide burden of cancer were provided by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO). passed away on September 12, 2018, in Geneva, Switzerland, from cancer. The database may be viewed online and is open to the public. Estimates of cancer incidence and death were the basis for the development of the IARC Global Cancer Observatory database. acquired from a variety of sources for 36 distinct cancers across 185 nations.

A number of reasons, including an aging population and a rise in the overall cancer burden, are contributing to this trend, as are changes in the incidence of certain cancers associated with social and economic progress. Cancers linked to industrialized-country lifestyles are replacing those linked to poverty and infections, especially in rapidly developing nations.

In 2018, the majority of new cases and fatalities from cancer were anticipated to occur in Asia. This is partly due to the fact that the area is home to roughly 60% of the world's population. The incidence of breast cancer has climbed by 39.1 percent from 1990 to 2016, surpassing all other cancers among women in India. Despite having just 9.0% of the world's population, Europe is responsible for 20.3% of cancer fatalities and 23.4% of all cancer incidences. The Americas are home to 13.3% of the world's population and produce 21.0% of the world's new cases and 14.4% of the world's deaths. Since certain cancer types in Asia and Africa are more common and are associ-

ated with worse prognoses and higher mortality rates, and because many countries in these regions lack access to timely diagnosis and treatment, the proportion of cancer deaths in these regions is 57.3% and 7.3%, respectively, rather than the proportion of incident cases, which is 48.4% and 5.8%, respectively. The most common cancers in women are those that affect the lungs, breasts, and colorectum.(7)

The bio-genetic-molecular process allows cancer to thrive and spread from one host organ to another. In 2018, cancer claimed the lives of 9.6 million people, surpassing the combined tolls of HIV/AIDS, malaria, and TB. Low- and middle-income countries (LMICs) are projected to account for the majority of the twofold increase in cancer incidence by 2035. Low- and middle-income nations account for 60% of all cancer cases and 75% of all cancer deaths.

New diagnostic and therapeutic technology avenues are emerging as a result of the integration of many research areas fostered by cancer R&D, including genetics, genomics, nanotechnology, nanomedicine, computer sciences, etc. Radiologists, cancer specialists, surgeons, and other respiratory doctors and nurses are all part of the interdisciplinary teams that work in every hospital throughout the globe. expert witnesses. In order to determine whether a patient has lung cancer, this interdisciplinary team analyzes a broad range of imaging tests, including computed tomography (CT), x-rays, PET (positron emission tomography), and biopsies. When a patient has these tests, the medical staff is better able to make a complete diagnosis. From the first signs of cancer to the most advanced tumors, medical experts utilize TNM categorization to better understand lung cancer. When cancer is in its initial stages, like IA, it's more likely to have been found by mistake, doing a biopsy becomes more challenging, and the prognosis improves (probability to live). Stage IV cancer, on the other hand, is poorer prognosis-wise, very symptomatic, and easy to identify (with biopsy).

Due to the tumors' microscopic size (less than 4 mm), medical practitioners have a hard time diagnosing them using CT scan pictures and making accurate predictions about their cancer causes or recurrence risks when dealing with IA classification. Although a biopsy is necessary, it may be too intrusive and impractical to do due to the 50/50 probability that IA tumors are benign or malignant. The difference between a benign tumor and a malignant one is whether the tumor spreads or not. In order to detect any signs of growth—which may increase in size by more than 100%—general medical recommendation is to have another scan in 5-10 weeks.eight (8)

The global human population faces a grave danger from the following hypothetical situation: "What if a computer is trained using unlabeled data? The algorithm can then be left to discover patterns in the data and potentially teach us new things." For example, in the case of extremely small tumors (less than 4 mm) that are difficult to diagnose using CT scan images, these algorithms would be invaluable to doctors and experts when human eyes fail to detect the desired patterns. The use of supervised and unsupervised machine learning methodologies in AI makes this a real possibility.two (2) One of the most common cancers in many nations and a top killer of men and women alike is lung cancer. Smoking, secondhand smoke, air pollution, and other societal risk factors are associated with lung cancer. No small cell lung cancer accounts for around 80% of lung cancer instances, whereas small cell lung cancer accounts for the other 20%. The five-year survival rate for lung cancer is dismal, and the death rate is quite high. one to ten percent of patients. In order to choose the right treatment—conventional chemotherapies, target therapies like gefitinib or erlotinib, or new immune therapies like nivolumab or atezolizumab—it is crucial to identify the type of lung cancer (e.g., adenocarcinoma or squamous cell carcinoma), mutations (e.g., epidermal growth factor receptor or anaplastic lymphoma receptor tyrosine kinase rearrangements), and sequential mutations (T790). Patients' chances of survival are impacted by this crucial procedure. Because standard chemotherapy and, more lately, targeted treatments are different for lung adenocarcinoma and lung squamous cell carcinomas, it is essential to diagnose lung cancer by classifying and mutating the kind of cancer. Applying molecular biomarkers to lung biopsies and/or blood tests allows for the correct identification of lung cancer type and stage, which in turn allows for the selection of treatment choices. Some negative impacts are caused by this technology as it is. The time spent with cancer is a critical variable for survival and, when possible, for healing; on the one hand, patients must undergo invasive surgical procedures for tissue biopsies or liquid biopsies (e.g., blood) and wait a certain amount of time (approximately one month) before they can receive a diagnosis of cancer type and the appropriate treatments. In contrast, hospitals incur a hefty expense when using these methods, which need specialized staff and the right tools—both of which are in short supply or overworked in some areas due to cutbacks in public lab and hospital funding. See if we can find a way to use AI engineering to solve these problems. two (2) Lung cancer classification using AI (Machine Learning & Deep Learning): Recent advances in cheap, powerful AI have enabled the training of larger and more complicated neural networks, and some studies have shown that deep learning can classify breast, bladder, and lung tumors.This has led to the development of multiple deep convolutional neural networks (CNNs) that can complete complicated visual recognition tasks, proving that deep learning technology can improve medical diagnosis and treatment. CNNs can achieve accuracies in cancer classification that are on par with pathologists'.

When evaluated on separate datasets of frozen samples and pictures derived from biopsies, these models retain their performance. Classification of lung patterns on computed tomography (CT) images has proved successful when CNN's ideas are utilized. Using this CNNs technology, Google Incorporated won the 2014 ImageNet Large-Scale Visual Recognition Challenge. For studies using visual object identification software, researchers have turned to the massive ImageNet project. The project has manually annotated over 14 million photographs to show what things are in them. Additionally, bounding boxes have been supplied for at least one million of these images, which is a subset of Unsupervised Machine Learning. By using microarchitecture units known as inception, ImageNet enhanced the GoogleNet architecture's translation robustness and nonlinear learning capabilities. A number of nonlinear convolution modules with varying degrees of resolution are included in each inception unit. In medicine, and especially in pathology, the ability to analyze data at numerous resolutions is a key component of the Inception architecture. A deep-learning model for the automated analysis of tumor slides using publicly available whole-slide images in The Can-

cer Genome Atlas (TCGA) has been developed by this complex Convolutional Neural Network (ANN), which has previously been successfully applied to specific disease classifications such as diabetic retinopathy detection and skin cancers. Researchers programmed inception v3 to distinguish between benign and malignant lung cancers. Later on, they put these techniques to the test by trying to differentiate between lung squamous cell carcinoma and adenocarcinoma, two more difficult types of lung cancer. The deep-learning model was also tested and trained using a straight three-way categorization into the three picture kinds. Three pathologists, including two thoracic pathologists and one anatomic pathologist, were evaluated for their ability to visually analyze and categorize the whole-slide pictures in the test set without relying on the classification supplied by TCGA. The results of this assessment were compared to those of the evaluation. In general, deep learning models performed as well as any individual pathologist. To save time and money, pathologists may now use new AI technology based on deep-learning models to aid in the diagnosis of cancer subtype or gene alterations in any form of cancer. What's more, these AI technologies can help underprivileged places can also have high benefits by sending the digital images to labs of other countries, generating a reduction of current gap in healthcare between countries. Machines are very helpful not only to identify but also classify, detect or distinguish tumors and other malignancies. In other words, machines have been used primarily as an aid to cancer diagnosis and detection useful for mankind. Artificial Intelligence (Machine Learning & Deep Learning) detecting the type of Breast cancer: Thousands of our grandmothers, mothers, and daughters fall victim to breast cancer every year. The human body comprises of millions of cells each with its own unique function. When there is unregulated growth of any of these cells it is termed as Cancer. In this, cells divide and grow uncontrollably, forming an abnormal mass of tissue called as tumor. Tumor cells grow and invade digestive, nervous and circulatory systems. Breast cancer is the most frequent cancer worldwide among women. Studies based on advanced countries show that the incidence of breast cancer has been increasing. It has been postulated that breast cancer incidence tends to be higher in more developed countries due to delayed Childbearing, revealed by (WHO), higher usage of hormone replacement therapy, a higher rate of screening, and improved tumor registries. Some studies have also argued that higher-income countries may have higher fat diets, consumption of processed foods, and an increased rate of obesity, both correlated with higher breast cancer incidence rates. In general, researches note that many Western populations have a higher incidence rate of breast cancer than regions in Africa and Asia Breast biopsies, like all types of cancer, are used to diagnose breast cancer type, stage, and mutations. An accurate breast cancer staging is an essential task performed by pathologists to inform clinical management. For instance, evaluation of the extent of cancer spread by histopathological analysis of sentinel axillary lymph nodes is an important part of breast cancer staging. The sensitivity of sentinel axillary lymph nodes evaluation by pathologists, however, is sub-optimal because some studies show that pathology review by experts changed the nodal status in 24% of patients this effect is due to the difficulty of the process of decision making of human behavior. Moreover, sentinel axillary lymph nodes evaluation is time-consuming and has a certain cost. In short, the high cost of exams and a period of more than one month to have breast cancer type, stage, etc. to decide the appropriate therapy, it can affect the results of therapeutic treatments and survival of patients. The application of Artificial Intelligence (AI) machine learning technology with deep learning algorithms to whole-slide pathology images can potentially improve the diagnostic accuracy of breast cancer and metastases. Some scholars have assessed the performance of automated deep learning algorithms at detecting metastases in tissue sections of lymph nodes of women with breast cancer results are accurate than the results compared with pathologists diagnoses, some deep learning algorithms achieved better diagnostic performance than a panel of 11 pathologists participating in a simulation exercise designed to mimic routine pathology workflow; in short, algorithm performance was comparable with an expert pathologist interpreting whole-slide images without time constraints. Hence, this experiment has shown that deep learning algorithms could identify metastases in sentinel axillary lymph nodes slides with 100% sensitivity, whereas 40% of the slides without metastases could be identified as such. This approach can significantly reduce the workload of pathologists and improve the management decisions on whether or not to administer therapy, perform a surgical intervention, etc. Overall, then, this interesting result shows the potential of AI applied in cancer imaging for the detection of lymph node metastases in women with breast cancer to improve processes of diagnosis and therapy of this and other critical cancers.(2)

The continuous development of technology in the medical field will save countless lives and the overall quality of human life continues to improve over time.

II. METHODOLOGY

A regression analysis looks for connections between different traits or variables. When it comes to statistics and machine learning, regression analysis is a major player. Data scientists use one of several available regression algorithms. Logistic Regression is the topic of this article.

For the purpose of making probability predictions about categorical dependent variables, Logistic Regression is a Machine Learning classification approach. A binary variable with values 1 (yes, success, etc.) or 0 (no, failure, etc.) indicates the dependent variable in logistic regression. Put simply, the logistic regression model uses X to predict the probability of Y=1.A binary dependent variable is necessary for binary logistic regression.To get a binary

Ideally, the dependent variable's factor level 1 would reflect the intended result of the regression.You should only include the relevant variables.It is important that the independent variables remain separate from one another. In other words, multicollinearity should be minimal in the model.The log chances have a linear relationship with the independent variables.Large sample sizes are necessary for logistic regression.For example, if a person has cancer, then the likelihood of developing diabetes, coronary heart disease, or cancer is 1. If he does not have cancer, then the likelihood of developing these diseases is 0. This is because logistic regression deals with discrete data, such as a person's age, sex, body mass index, and the results of different blood tests. One item that might be removed from the list

is the Patient ID feature since it does not provide any useful information for cancer prediction. It is necessary to construct a logistic regression model in order to regress the independent features (X-axis) against the dependent feature (Y-axis), which is the target (Y-axis). Features that are deemed independent include lump thickness, uniformity of cell size and shape, marginal adhesion, size of single epithelial cell, bare nuclei, bland chromatin, normal nucleoli, and mitoses, whereas feature target is considered dependent.

With the use of the publicly accessible Wisconsin data set and Anaconda Python 3.6, we were able to create a Logistic Regression model, which achieved an accuracy rate of 96.5%.

Features Explanation:

Now we need to Identify for every Patient ID, how every Independent features (X-axis) are closely correlating to the Output or Responses or Dependent Variable (Y-axis), what should be the Diagnosis for every patient ID falls under either. Diagnosis (M being malignant, B being benign). Which are considered as Y – axis or Target variable.

III. RESULTS

The breast cancer data with 10 attributes was analyzed to identify the error rates and accuracy of 649 patients. The accuracy of the model is 96.56%, which shows 96.5% is correct predictions.

IV. DISCUSSION

Predicting whether a patient's tumor is malignant or benign is the goal of this logistic regression.We are saddened by the fact that breast cancer is the leading cause of death among women globally; over half a million women lost their lives to this disease in 2018. Another estimate shows that out of 500,000 breast cancer patients, 90,000 are misdiagnosed; unfortunately, this is a significant problem that requires the use of these cutting-edge methods to eradicate. I think there's a way to use Machine Learning to make breast cancer diagnoses more efficient; the way they're done now is inefficient.Not many Guidelines for the most effective detection of breast cancer via screening Breast cancer screenings: A mammography is the gold standard for breast cancer screenings. Breast X-rays are called mammograms. Up to two years before either you or your doctor would notice, it may identify breast cancer.Women in their forties and fifties who are at ordinary risk of

annually should have a mammogram to detect breast cancer. Beginning at age 30, high-risk women should get annual mammograms in addition to magnetic resonance imaging (MRI).

There are a lot of things that might put you at risk for breast cancer. So far, we have not been able to pinpoint a single cause of breast cancer.

Age: The likelihood of developing breast cancer rises with the passage of time. The majority of breast cancer patients are older women, accounting for over 80% of cases.

One increased risk of acquiring cancer in the other breast is associated with a personal history of breast cancer, which is more common among women who have already had the disease in one breast.

If a patient's mother, sister, or other blood-related female has had breast cancer, particularly in her youth (before 40), the patient's chance of developing the disease is increased.

Heredity: Some females have certain mutations in their DNA. For instance, alterations to the BRCA1 and BRCA2 genes. Breast cancer is more likely to occur in people with certain genetic abnormalities. Breast cancer risk factors may also be increased by other genetic alterations.

Period history: the likelihood of developing breast cancer increases when a woman reaches her first childbearing years older.Women who menstruate for the first time before 12. A woman's menstrual period begins once she becomes 55 years old. Women who haven't had children. (5)

REFERENCES

[1] The Third Edition of Artificial Intelligence John Henry Winston, Jr.

[2]A Revolution in Cancer Treatment with AI: A New Standard in Technology Penelope De Lucchia

the thirdNational Institutes of Health, 2019. May 2019, International Agency for Research on Cancer

[4] https://www.sciencedirect.com/search/advanced?qs=artificial%20intelligence%20and%20retinopathy&show=25&sortBy=relevance&lastSelectedFacet=years&years=1998. (2019, Advanced Research). as of May 2019

[5]s://medium.com/@nyla.pirani/how-to-build-a-basic breast cancer model machine learning-9bae53c40f18, https://realpython.com/linear-regression-in-python/

[6] Machine Learning's Role in Cancer Diagnosis and Forecasting In the fields of biology and prediction, Joseph A. Cruz and David S. Wishart work in the departments. David A. Cruz, Joseph A. Cruz P. Wishart

The sentence is paraphrased as: "Search results for the term "global health organization cancer" will be returned with the parameters "oq" and "sourceid" set to "chrome."

[8] Deep Convolutional Networks for Detecting Lung Cancer-Salamon Jelo