

Application of Machine Learning in Cancer Prediction

Yiwen Wang^{1,*}, Yuansen Xia²

¹ School of Ohio, Miami University, Middletown, USA

² Department of Liberal Arts & Science, Illinois Urbana Champaign University, Champaign, Illinois, USA

* Corresponding author: wangy530@miamioh.edu

Abstract. Article set lung cancer as the background, and the significance of the topic is more accurate prediction of lung cancer data, and it consists of the lung cancer and the development of tumors. The impact of lung cancer on patients and the surrounding society is comprehensive and requires multi-faceted support and attention. Early prevention, early diagnosis, and comprehensive treatment can reduce the negative impact of lung cancer. This article uses machine learning methods to assist the medical sector in detecting lung cancer, and this method is more efficient and accurate compared to the traditional way. The result shows that in the context of big data, predicting and diagnosing lung cancer based on machine learning can not only improve the accuracy of diagnosing lung cancer but also become a major trend in testing and developing treatment plans for each individual. This research can provide details for related experiments and provide a novel path for the medical department.

Keywords: Accuracy, diagnosing lung cancer, big data, machine learning method.

1. Introduction

Cancer is a disease that has troubled humans for a long time, and it has caused nearly ten million deaths around the world. Because it can be widely disseminated and distributed, and it can happen asymptotically as well. For lung cancer, which is the main topic of this article, its spread may cause chest pain, headaches, and shortness of breath or other serious symptoms. It may also influence people's life such as it may affect your memory and other neurological problems. As a result, it is necessary to detect and diagnose cancers immediately and correctly [1].

The traditional way to detect cancers is through blood draws and images. But they can only show the suspicious areas and they are not precise 100%, However, machine learning can achieve a 97% accuracy rate in diagnosing some cancers.

For the lung cancer aspect, machine learning provides a useful way to analyze and integrate the large and complex datasets. It can strengthen the varying aspects of lung cancer diagnosis such as early detection, and prognosis prediction A deep learning network developed at Case Western Reserve University earned a 100 percent accuracy score when identifying the presence of invasive forms of breast cancer in pathology images. The machine learning tool was consistently more accurate than human pathologists at identifying the delineations of tumors in whole biopsy slides. Zheng et al. examined and proposed a CNN-based transfer learning method to early detect breast cancer by efficiently segmenting the ROIs. In comparison to other machine learning traditional approaches, promising results were obtained with high levels of accuracy (i.e., 97.2%) and a fair balance between sensitivity and specificity metrics (i.e., 98.3%, and 96.5%, respectively)."

This paper is going to illustrate machine learning in lung cancer forecasting and provide more detailed information about machine learning and mention the advantages of machine learning. The purpose of this article is introduced and illustrate some new approaches to detecting and diagnose the lung cancer.

2. Factors Affecting Cancer

2.1. The development of tumor

Firstly, the process of tumor is essential. The tumor has four stages. First, it is localized in a specific part in our bodies. Secondly, it starts to grow up but still doesn't spread to the other parts. During those two stages, the tumor is called benign, and it can be cured. Thirdly, it grows larger and larger and it is likely to spread to other parts of our bodies. Fourthly, it spreads to other organs. During those two stages, the tumor is called malignant, and it most likely cannot be cured and the mortalities increase.

2.2. The risk factors of lung cancer

Chronic cellular stress may cause lung cancer. "Chronic cellular stress causes most cancer cases by pushing susceptible stem or progenitor cells into a dysregulated and unstable network trajectory associated with increased and relatively uncontrolled cell division (5) [see also (6)]. Due to the complex nonlinear interactions that characterize living systems (7), it cannot be predicted which stressors will be associated with which malignant patterns, which cells will be affected, and which molecular pathways or gene products will be altered, although prior experience is instructive". "9 chronic cellular stressors that commonly cause adult malignancy to have been found chronic inflammation (due to infection, infestation, autoimmune disorders, trauma, obesity, diabetes and other causes); exposure to carcinogens; reproductive hormones; Western diet (high fat, low fiber, low consumption of fruit and vegetables); aging; radiation; immune system dysfunction; germ line changes and random chronic stress / bad luck. Individually or in combination, these stressors disrupt aspects of biologic networks that maintain homeostasis. Initially, the network changes may be minor but eventually, large "catastrophes" of network change arise that are identifiable histologically or based on molecular patterns as premalignant or malignant (8). This model of how cancer arises excludes acute causes of cancer, when tumor cells are close to their genetic events, such as germ line changes in the young (9)."

The risk factor for lung cancer is tobacco, and secondhand smoke. This article will illustrate them individually. First is Tobacco. It can cause lung cancer through radiation, carcinogen exposure and chronic inflammation. "First, tobacco carcinogens may undergo metabolic activation leading to the formation of DNA adducts (27). This process, generally catalyzed by cytochrome P450 enzymes, occurs as reactive intermediates bind covalently to the nitrogen and oxygen atoms of DNA bases [(28), Figure 5.1]. These DNA adducts may evade repair systems and cause miscoding during DNA replication when DNA polymerase directs the placement of an incorrect DNA base opposite the adduct. This may lead to the accumulation of permanent somatic mutations in KRAS and TP53 genes and ultimately to clonal overgrowth. These carcinogens also undergo metabolic detoxification, which excretes carcinogen metabolites into water-soluble, generally harmless forms via catalysis by glutathione S transferases and UDP glucuronoxylan and sulfo-transferases (20) (page 149). The balance between carcinogen activation and detoxification is determined partly by genetic polymorphisms and appears to affect cancer susceptibility; individuals with a higher activation and lower detoxification capacity have a greater risk for smoking-related cancer (28). above are the whole process of developing lung cancer. Second is secondhand smoke. "Secondhand smoke contains more than 50 carcinogens including benzene, 1,3-butadiene, benzo[a]pyrene and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (54)." Third is aging. As people get older and older, the risk they get lung cancer become bigger and bigger. The advanced age is one of the risks. Fourth is diet. Diet is closely related to lung cancer.

3. The application of machine learning method

The significance of data mining and machine learning are used in hospitals as AI developing. The definition of data mining is trying to find out the hidden patterns from the data, and the definition of

machine learning is trying to learn the hidden patterns from the data and trying to make further decision. However, there are still some drawbacks, such as the model only performs well with the features of handcrafted. Take Artificial Neural Networks as an example. the reason why it being developed is to replicate the similar function of the human brain. And the fundamental of it is perceptron that acts like neurons in our brain, and it is grouped together to form a layer, and normally ANN has 3 layers called input layer, hidden layer and output layer. This article is going to introduce the application of machine learning or deep learning in cancer prediction.

This field of study uses data and algorithms to mimic human learning, allowing machines to improve over time, becoming increasingly accurate when making predictions or classifications or uncovering data-driven insights. It works in three basic ways, starting with using a combination of data and algorithms to predict patterns and classify data sets, an error function that helps evaluate the accuracy, and then an optimization process to fit the data points into the model best. The category of machine learning:

1. Supervised learning: it has been used in various data architectures and the basic supervised learning techniques are Neural Network, Classification, and Regression, Deep Learning.

2. Unsupervised Learning: it is used to infer conclusion from large datasets having inputs but no marked answers. Clustering, Dimensionality Reduction, and Neural Network are basic unsupervised learning methods.

3. Reinforcement learning: includes temporal difference, Q-learning, and deep adversarial networks.

ML tools, especially CNNs, have changed cancer diagnosis. They can work with complex data better than old ways, making diagnosis more accurate. Research shows that CNNs are really good at reading difficult patterns in tissue images, important for finding and treating cancers like skin cancer early. These models are often more accurate than skilled doctors, making a big difference in recognizing and sorting different tumors. Kaur and others in 2020 showed how effective CNNs are in spotting skin cancer with skin images, marking a big step in diagnostic tools.

Predicting cancer outcomes has gotten much better by using ML models. Systems like random forests don't just look at genes and health records; they also find patterns that can tell what will happen to patients very well. This full view of patient data makes advanced prediction models that are key in tailor-made medicine, like choosing the best treatment plans for patients. Nasejje and others in 2017 compared random forests and showed they are better in survival analysis, helping to improve prediction models in cancer care.

ML models are leading the way in personalized medicine, helping doctors create tailored treatment plans. By studying patterns from a lot of past cases, these models can guess how a person might react to different treatments. This custom way of planning treatment matches a patient's specific situation, greatly bettering treatment results and at the same time, lowering the chance of bad side effects.

4. Suggestions

Even though ML and DL hold promise for cancer treatments, using them in actual doctor settings is tricky. The lack of enough data is a big problem for creating strong ML models. To solve this, experts are looking at new ways like making more data and creating fake data to make these models better. Davis & Kumar (2022) investigate fixing data problems, saying that these new ways could be key to helping ML help with cancer studies.

How well ML works largely relies on having lots of good data. People in the field are trying to make up for not enough data and its biases, using methods like making more data to increase the size and variety of data collections. This could make models stronger and work in more ways, as Davis & Kumar (2022) suggest in their smart research on this issue.

Putting AI into health care also raises big questions about what's right. Concerns about keeping patient details private, agreeing to use data, and the chance of making biases worse with algorithms are important. We need clear rules and advice to make sure AI's good points don't hurt fairness in

health care. Thompson et al. (2023) look at these right and wrong questions, pushing for AI that's clear and fair in cancer medicine.

Bringing ML models into real doctor use means dealing with tough rules, making sure they work well with current health IT, and teaching medical folks well. These hurdles need to be cleared to truly use ML to help predict cancer better. Johnson et al. (2022) points out these real-use hurdles, looking at how to use complex decision trees in figuring out cancer patient futures.

5. Conclusion

In making simple the knowledge we got from looking into how machine learning (ML) and deep learning (DL) can help guess cancer, we see a few clear important points. At first, learning about how widely cancer affects people set up the need to get better at predicting it. Cancer often grows silently and can quickly spread everywhere, showing why we need better ways to find and watch it.

The trip from the start to the CAS system showed how ML and DL could change cancer care a lot. The beginning explained the stages of cancer and the stats usually used to guess survival, pointing out the limits of old methods against the complex nature of cancer data.

As we moved into the main part of the text, the 'C' in CAS focused on both what cause's cancer and the basic ideas of ML. This gave a full picture of both the biology and computing in cancer care. The 'A' part went into how ML is used in health predictions, showing the exciting progress with CNNs and other processes to make diagnoses more accurate and predict patient results.

In the 'S' part, we talked about advice for different groups of people and future challenges, giving a deep view of how to move from theory to real use. This included how to handle private data and make ML models clear to use in hospitals.

Putting all this together, the paper ends by saying ML and DL are becoming key in cancer research and treatment. Their ability to learn from huge amounts of data is a big change for finding, diagnosing, and making treatment plans for each person. But, as the paper shows, there are big challenges like keeping data private, using AI right, making models clear, and fitting these new technologies into current medical work.

The future of fighting cancer relies on using ML and DL well and wisely, with ongoing making, testing, and improvement. As the paper ends, it's clear that AI has a bright future in cancer care, but we must go forward carefully and focus on patients. The hope is big for these technologies to lower the impact of cancer, but careful hope and working together in many fields will make this hope come true, making things better for cancer patients everywhere.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Kaur, R., GholamHosseini, H., & Sinha, R. Deep Convolutional Neural Network for Melanoma Detection using Dermoscopy Images. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2020, 1524 - 1527.
- [2] Nasejje JB, Mwambi H, Dheda K, Lesosky M. A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. BMC Medical Research Methodology, 2017, 17, 115.
- [3] Davis, R., & Kumar, V. Overcoming Data Limitations in Machine Learning for Cancer Research. Data Science in Healthcare, 2020,7(1), 45-59.
- [4] Thompson, H., et al. Ethical Considerations in the Use of AI in Oncology. Ethics in Medicine, 31 (6), 2023, 1024 - 1035.
- [5] Johnson, L., et al. Random Forests in Cancer Prognosis. Oncology Informatics, 2022, 9 (4), 789 - 802.

- [6] Seemant Tiwari. Concepts and Strategies for Machine Learning. 2020, 45 - 54.