International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

A review of literature on diagnosis of cancer using deep learning techniques

Sagar K M[1], Sachitha M V[2], Shreyas K[3]

[1,2,3]*Dept. of Information Science and Engineering, MIT Mysore*

---***---

Abstract - *Cancer has a high death rate because of its aggressiveness, enormous metastatic potential, and heterogeneity (which results in chemotherapeutic resistance). One of the most common types of cancer that can affect both sexes and occur worldwide is lung and colon cancer. The quality of treatment and the survival rate of cancer patients can both be significantly increased by an early and precise diagnosis of these cancers. 25000 histopathological photos of lung and colon tissues, evenly divided into 5 groups, make up the dataset. The simultaneous integration of data from a wide variety of scales, from nuclear aberrations (O(0.1 m)) through cellular structures (O(10 m)) to the overall tissue architecture, is essential for histopathologic diagnosis. Normal cell growth follows a predictable pattern. When cancer develops, a collection of cells suddenly begins to expand haphazardly and uncontrollably, resulting in lumps or tumors. A malignant tumor can spread to other areas of the body and never cease growing. A high-fat, low-fiber diet is associated to colorectal cancer. It is a disease of the wealthy and is more common in wealthy nations. This cancer can cause diarrhea, abdominal, lower back, or bladder pain, as well as changes in bowel habits. Smoking causes 90% of lung cancer cases. Lung cancer risk has grown due to pollution in urban areas. A chronic cough, weight loss, and appetite loss are possible symptoms.*

Key Words: Histopathological images, nuclear aberrations, tumor.

1. INTRODUCTION

Cancer is the unchecked expansion of aberrant cells anywhere in the body. Cancer cells, malignant cells, or tumour cells are names for these aberrant cells. These cells are able to invade healthy body tissues. The names of the tissues from which the aberrant cells originated (such as breast, lung, and colorectal cancer) help further identify many cancers and the abnormal cells that make up the cancer tissue. A mass of cancer cells forms when damaged or unrepaired cells continue to divide and proliferate uncontrollably without dying and turning into cancer cells. Cancer cells frequently separate from this initial clump of cells, move through the blood and lymphatic systems, and settle in other organs where they can restart the unchecked development cycle. Metastatic spread or metastasis refers to the process through which cancer cells leave one part of the body and spread to another. For instance, if breast cancer cells have metastasized to the bone, the patient has this condition. Any factor that could lead to an aberrant body cell's development is possibly carcinogenic.

Some cancers have unidentified origins, whilst others have environmental or lifestyle triggers or may have many known causes. Some traits may be altered by a person's genetic make-up during development. Combinations of these causes frequently result in cancer development in patients. The following malignancies have been specifically related to human genes: breast, ovarian, colorectal, prostate, skin, and melanoma. The likelihood that someone may acquire cancer increases with the quantity or level of cancer-causing substances to which they are exposed. Additionally, those who are genetically predisposed to cancer may not experience it for identical reasons (lack of sufficient stimulation to activate the genes). Additionally, some people may have an immune response that is more active than usual, controlling or eliminating cells that are or may develop into cancer cells. Carcinogens are chemicals that have the potential to mutate organisms. Benzene, asbestos, nickel, cadmium, vinyl chloride, benzidine, N-nitrosamines, tobacco, and cigarette smoke (which contain at least 66 recognized potentially carcinogenic compounds and poisons) are examples of carcinogens.

2. LITERATURE REVIEW

Dataset

Nuclei instance segmentation plays an important role in the analysis of Hematoxylin and Eosin (H&E)-stained images. While supervised deep learning (DL)-based approaches represent the state-of-the-art in automatic nuclei instance segmentation, annotated datasets are required to train these models. There are two main types of tissue processing protocols, namely formalin-fixed paraffin-embedded samples (FFPE) and frozen tissue samples (FS). Although FFPE-derived H&E stained tissue sections are the most widely used samples, H&E staining on frozen sections derived from FS samples is a relevant method in intra-operative surgical sessions as it can be performed fast. Due to differences in the protocols of these two types of samples, the derived images and in particular the nuclei appearance may be different in the acquired whole slide images. Analysis of FS-derived H&E stained images can be more challenging as rapid preparation, staining, and scanning of FS sections may lead to deterioration in image quality[1].

The National Lung Screening Trial (NLST) was a randomized controlled trial conducted by the Lung Screening Study group (LSS) and the American College of Radiology Imaging Network (ACRIN) to determine whether screening for lung cancer with low-dose helical computed tomography (CT) reduces mortality from lung cancer in high-risk individuals relative to screening with chest radiography. Approximately 54,000 participants were enrolled between August 2002 and April 2004. Data collection has ended, and information is complete through December 31, 2009[2].

Developing a tissue bank database has become more than just logically arranging data in tables combined with a search engine. Current demand for high quality samples and data, and the ever-changing legal and ethical regulations mean that the application must reflect TuBaFrost rules and protocols for the collection, exchange and use of tissue. To ensure continuation and extension of the TuBaFrost European tissue bank, the custodianship of the samples, and hence the decision over whether to issue samples to requestors, remains with the local collecting center. The database application described in this article has been developed to facilitate this open structure virtual tissue bank model serving a large group. It encompasses many key tasks, without the requirement for personnel, hence minimizing operational costs[3].

The LUNA16 (LUng Nodule Analysis) dataset is a dataset for lung segmentation. It consists of 1,186 lung nodules annotated in 888 CT scans. Source: Universal Lesion Detection by Learning from Multiple Heterogeneously Labeled Datasets[4]. Dataset of histopathological images of lung and colon cancer called LC25000 from the publicly available Kaggle website to evaluate the proposed systems. The dataset was compiled by Andrew Borkowski and his associates at James Hospital Tampa, Florida, which consists of 25,000 images divided into two types of colon cancer and three types of lung cancer. The images are distributed among the five types equally, meaning the dataset is balanced, and each type contains 5000 images. These types are colon_aca (Adenocarcinoma) and colon_bnt (Benign Tissue), lung_aca (Adenocarcinoma), lung_bnt (Benign Tissue), and lung_scc (Squamous Cell Carcinoma). Colon adenocarcinoma accounts for more than 95% of colon cancers due to the non-removal of polyps in the large intestine. Adenocarcinoma of the lung accounts for more than 40% of lung cancers, which appear in glandular cells and spread within the lung and alveoli. Lung squamous cell carcinoma accounts for more than 30% and is the second most common type of lung cancer, which appears in the bronchi. The other two types are benign and do not spread to other parts of the body. However, it must be effectively verified by biopsy and removal[5][8][9][11].

PatchCamelyon is an image classification dataset. It consists of 327.680 color images (96 x 96px) extracted from histopathologic scans of lymph node sections. Each image is annotated with a binary label indicating presence of metastatic tissue. PCam provides a new benchmark for machine learning models: bigger than CIFAR10, smaller than ImageNet, trainable on a single GPU[7].

2.2 Methodology

SegNet is a semantic segmentation model. This core trainable segmentation architecture consists of an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. The architecture of the encoder network is topologically identical to the 13 convolutional layers in the VGG16 network[1]. UNET is an architecture developed by Olaf Ronneberger et al. for Biomedical Image Segmentation in 2015 at the University of Freiburg, Germany. It is one of the most popularly used approaches in any semantic segmentation task today. It is a fully convolutional neural network that is designed to learn from fewer training samples. It is an improvement over the existing FCN — “Fully convolutional networks for semantic segmentation” developed by Jonathan Long et al. in 2014[1][6].

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image[2][3][4][5][9][11][13].

A capsule neural network (CapsNet) is a machine learning system that is a type of artificial neural network (ANN) that can be used to better model hierarchical relationships. The approach is an attempt to more closely mimic biological neural organization. The idea is to add structures called "capsules" to a convolutional neural network (CNN), and to reuse output from several of those capsules to form more stable (with respect to various perturbations) representations for higher capsules. The output is a vector consisting of the probability of an observation, and a pose for that observation. This vector is similar to what is done for example when doing classification with localization in CNNs[8][15].

3. RESEARCH GAPS

Comparison with different models yield in better understanding of the dataset. The LCP-CNN has been trained to generate desired output on lung cancer dataset. In this literature survey, the use of Tensor flow backend with GPU to speed up a machine learning algorithm.

Using this computer-based identification method in the medical centers will allow pathologists to diagnose more lung and colon cancer cases in less effort, cost and time. Multi-scale multi-encoder models improve histopathology image segmentation. Performance metrics are compared with existing deep learning algorithms, such as CNN, deep neural network (DNN), LSTM, GRU and BiLSTM for binary tasks to demonstrate the efficiency of the proposed AlexNet-GRU. The proposed method improves CapsNet and can be adopted as a computer-aided diagnostic method to support doctors in lung and colon cancer diagnostics as Cancer. Improvement in computationally efficient and reduced time consumption is noticed from this work. The SSL method leverages both labeled and unlabeled data which are used to provide a low-cost alternative in terms of the requirement of the laborious and sometimes impractical sample labeling. Various Machine learning algorithm and technique were used like CNN, RNN, Alexnet, GRU etc. Various Methodology Like Alexnet GRU, Baseline architectures, msY model family, DFD-Net, DCNN, VGG, ResNet, Dense Net, Inception model, SegNet.

3. CONCLUSION

A plan for the diagnosis and treatment of cancer is a key component of any overall cancer control plan. Its main goal is to cure cancer patients or prolong their life considerably, ensuring a good quality of life. In order for a diagnosis and treatment programme to be effective, it must never be developed in isolation. It needs to be linked to an early detection programme so that cases are detected at an early stage, when treatment is more effective and there is a greater chance of cure. It also needs to be integrated with a palliative care programme, so that patients with advanced cancers, who can no longer benefit from treatment, will get adequate relief from their physical, psychosocial and spiritual suffering. Furthermore, programmes should include an awareness-raising component, to educate patients, family and community members about the cancer risk factors and the need for taking preventive measures to avoid developing cancer.

Analysing the results of lung and colon cancer requires specific medical information, including diagnostic tests, imaging studies, pathology reports, and individual patient history. For lung cancer, the results are typically analysed based on the type of cancer (such as adenocarcinoma or squamous cell carcinoma), the stage of the cancer (ranging from stage 0 to stage IV), and other factors like tumor size, lymph node involvement, and metastasis (spread to other parts of the body). This information helps determine the appropriate treatment options and prognosis. For colon cancer, the results are also evaluated based on factors such as the type of cancer (adenocarcinoma being the most common), the stage of the cancer, the presence of lymph node involvement, the presence of metastasis, and the characteristics of the tumor (such as tumor grade or differentiation). These details guide treatment decisions and provide insights into the outlook for the patient.

To fully understand and interpret the results of lung and colon cancer, it's essential to consult with a qualified healthcare professional or oncologist who can provide personalized analysis and guidance based on the specific circumstances of the individual patient. They will consider all relevant medical information and discuss the implications of the results, along with treatment options, prognosis, and any necessary follow-up or further testing.

REFERENCES

[1] Armitage P, Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. British Journal of Cancer 8:1–12.
[2] Barclay BJ, Kunz BA, Little JG, Haynes RH. 1982. Genetic and biochemical consequences of thymidylate stress. Canadian Journal of Biochemistry 60:172–184.
[3] Baron JA, Sandler RS, Haile RW, Mandel JS, Mott LA, Greenberg ER. 1998. Folate intake, alcohol consumption, cigarette smoking, and risk of colorectal adenomas. Journal of the National Cancer Institute 90:57–62.
[4] Bartek J, Bartkova J, Vojtesek B, Staskova Z, Rejthar A, Kovarik J, Lane DP. 1990. Patterns of expression of the p53 tumour suppressor in human breast tissues and tumours in situ and in vitro. International Journal of Cancer 46(5):839–844.
[5] Baylin SB, Herman JG, Graff JR, Vertino PM, Issa JP. 1998. Alterations in DNA methylation:A fundamental aspect of neoplasia. Advances in Cancer Research 72:141–196.
[6] Benn J, Schneider RJ. 1994. Hepatitis B virus HBx protein activates Ras-GTP complex formation and establishes a Ras, Raf, MAP kinase signaling cascade. Proceedings of the National Acad emy of Sciences 91:10350–10354.
[7] Bennett WP, Hollstein MC, Metcalf RA, Welsh JA, He A, Zhu S, Kusters I, Resau JH, Trump BF, Lane DP, Harris CC. 1992. p53 mutation and protein accumulation during multistage human esophageal carcinogenesis. Cancer Research 52:6092–6097.
[8] Bird A. 1992. The essentials of DNA methylation. Cell 70:5–8.
[9] Bohlke K, Cramer DW, Trichopoulos D, Mantzoros CS. 1998. Insulin-like growth factor-I in relation to premenopausal ductal carcinoma in situ of the breast. Epidemiology 9(5):570–573.
[10] Bohr VA, Anson RM. 1995. DNA damage, mutation and fine structure DNA repair in aging. Muta tional Research 338:25–34.
[11] Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps.
[12] An Efficient Deep Learning Approach for Colon Cancer Detection” by Ahmed S. Sakr, Naglaa F. Soliman , Mehdhar S.

Al-Gaashani , Paweł Pławiak, Abdelhamied A. Ateya and Mohamed Hammad in 2022.

[13] Lung and Colon Cancer Classification Using Medical Imaging : A Feature Engineering Approach by LARIS, SFR MATHSTIC, LaTIM, INSERM in 2021.

[14] Translational Oncology by Yue Xi, Pengfei Xu in 2021.

[15] Cancer Statistics by Rebecca L. Siegel, MPH ; Kimberly D. Miller, MPH ; Hannah E. Fuchs, BS; Ahmedin Jemal in 2021.

[16] Lung cancer detection by Bijaya Kumar et.al in 2020.

[17] Breast cancer classification from histopathological images using patch-based deep learning model by Irum Hirra et.al in 2020.

[18] Classification of histopathological images for early diagnosis of breast cancer by Ebrahim et.al in 2020.

[19] Transfer learning assisted multi-resolution breast cancer histopathological image classification by Nouman et.al in 2021.

[20] Breast cancer histopathological image classification based on deep feature fusion and enhanced routing by Pin Wang et.al in 2021.