A WEB APPLICATION USING MACHINE LEARNING MODEL TO DETECT CERVICAL CANCER IN EARLY STAGE

Sharmin Sultana

Department of Computer Science, University of Massachusetts Lowell

sharmin_sultana@student.uml.edu

ABSTRACT

Cancer has been classified as a heterogeneous disease with numerous subtypes. Among cancers, Cervical cancer remains one of the most prevalent gynecologic malignancies, worldwide. As cervical cancer is a highly preventable disease, therefore, early screening represents the most effective strategy to minimize the global burden of cervical cancer. Doctors use cancer imaging tests to evaluate if growth is malignant or not. If it is cancer, how quickly is it developing? What size is the spread? It had treatment; has it recovered? Studies suggest that AI could improve the responsiveness, accuracy, and dependability of medical personnel. AI has the potential to assist in earlier cancer diagnosis in patients who are already displaying symptoms. That's why we'd like to suggest a web-based tool by using Machine Learning to accurately predict cervical cancer stages from screening photographs.

KEYWORDS

Machine learning, Cancer susceptibility, Predictive models, Cancer recurrence, Cancer survival

ABBREVIATIONS

ML, Machine Learning; ANN, Artificial Neural Network; SVM, Support Vector Machine; DT, Decision Tree; BN, Bayesian Network

INTRODUCTION

Cancer research has undergone a continuous evolution over the last few decades [Hanahan and Weinberg (2011)]. Diagnosing cancer early on can improve a patient's treatment and prognosis. But detecting cancer in its first stage can be difficult, and current screening methods often require invasive procedures or expensive imaging equipment to identify the initial signs of disease. Scientists used various methods, such as early-stage screening, to detect cancer types before they cause symptoms. Furthermore, they have created new strategies for predicting cancer treatment outcomes early on. Large amounts of cancer data have been collected and made available to the medical research community as a result of the introduction of new technologies in the field of medicine. However, accurate disease prediction is one of the most

interesting and difficult tasks for physicians. As a result, machine learning methods have grown in popularity among medical researchers. These techniques can discover and identify patterns and relationships among them in complex datasets.

Cervical cancer (CC) is one of the leading causes of death for women in developing countries such as Bangladesh, India, and Pakistan [Olusola, Banerjee, Philley, and Dasgupta (2019)]. According to the American Cancer Society, nearly 13,800 women were newly diagnosed with invasive forms of CC in the United States in 2020, with 4290 deaths due to CC. Human papillomavirus (HPV) is a major cause of cervical cancer (CC), and CC incidence has been decreasing in countries with high HPV vaccine uptake[Olusola et al. (2019) and Burchell, Winer, de Sanjosé, and Franco (2006)] despite the fact that there are many factors associated with CC with lower levels of association. An important question is therefore whether CC can be predicted using a combination of patient data.

Because the prognosis is critical for long-term survival, early detection of breast cancer aids in early diagnosis and treatment. Because early detection, diagnosis, and treatment of cancer reduce the risk of death, it plays an important role in saving the patient's life. Any delay in detecting cancer in its early stages leads to disease progression and treatment complications; thus, a long waiting period between being diagnosed with breast cancer and beginning treatment is a cause for concern.

RELATED WORKS

In this section, we will briefly discuss the contributions of machine learning in early-stage cancer prediction.

Cancer Prediction and Prognosis

Sruthi et al. (2022) made a cancer prediction webpage using machine learning that includes three types of cancer: breast cancer, lung cancer, and prostate cancer and they use the SVM algorithm for breast cancer and the Random Forest algorithm for lung and prostate cancer. Prerita, Sindhwani, Rana, and Chaudhary (2021)'s goal is to review a Python methodology and its application in cancer diagnosis and prognosis by developing a basic machine-learning model like Logistic Regression, K-Neighbours Classifier, Support Vector Classifier linear, Gaussian Naive Bayes, and Decision Tree Classifier. SVM with polynomial kernel along with a 3rd degree provides above 99% accuracy for Breast Cancer Detection in microwave imaging [Sami, Sagheer, Riaz, Mehmood, and Zubair (2021)]. Unsupervised learning techniques could be used for predicting and detecting cancer tissues and provides merits of the proposed method over the existing cancer detection methods is the possibility of applying data from different types of cancer which describes the feature automatically and it helps to enhance the prediction and detection capabilities very specifically [Saxena and Prasad (2019)].

Early Stage Cancer Prediction

Dildar et al. (2021) reviews of deep learning techniques for the early detection of skin cancer. Bankar, Padamwar, and Jahagirdar (2020) aims to investigate the symptoms of the various age groups Youth, Working Class, and Elderly, and to calculate relative feature importances,

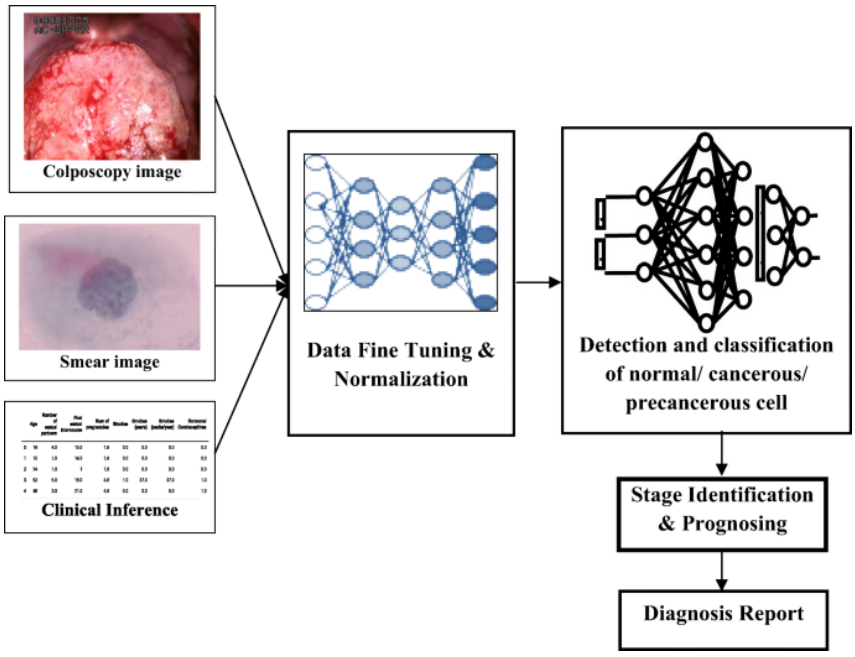

Figure 1. Early Stage Cervical Cancer Detection

tree-based algorithms such as Decision Trees, Random Forest, and XGBoost were used to identify the underlying data patterns. They concluded that Coughing Blood, Clubbing of Finger Nails, Genetic Risk, Passive Smoking, and Snoring are the factors that are responsible for lung cancer in all age groups in most cases. The diagnosis of cervical cancer is approached in []Moldovan (2020) using a machine learning method in which the features are selected using linear correlation and the data is classified using the Support Vector Machines (SVM) algorithm. The SVM hyperparameters are chosen using Chicken Swarm Optimization (CSO). The method is tested and validated using the UCI Machine Learning Repository's publicly available Cervical cancer (Risk Factors) Data Set. Early prediction of breast cancer using ANN highest accuracy for the activation functions used for forecasting namely Relu and Sigmoid [Mridha (2021)].

PROPOSED METHODOLOGY

In front end, users will upload their diagnosis images and after analyzing the images it will predict the stages of cervical cancer if detected. Figure 1 describes the back end operation. To begin, photos must be collected. Following that, the images will be preprocessed using a geometric mean filter. This eventually leads to an increase in image quality. The images will be then segmented using the K-means method. This segmentation makes it easier to identify the region of interest. Following that, machine learning-based categorization strategies are used. Figure 2 depicts the classification and prediction of cervical cancer using machine learning and image processing technology.

The pre-processing of images is important in the correct classification of photographs of ill-nesses. CT scans produce images that contain a wide range of artifacts, including noise, which can be seen in these scans. Image filtering techniques can be used to remove these artifacts. In order to reduce the amount of noise in the input images, a geometric mean filter will be applied.

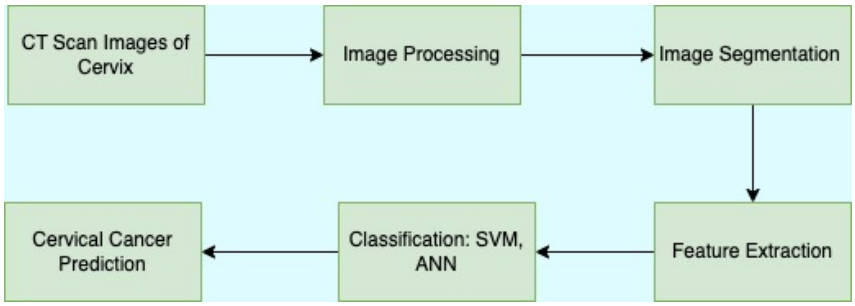

Figure 2. Flowchart of the Steps

After Segmentation, features will be extracted using LDA or PCA method then this data will be fed for classifications. We will use SVM and ANN to predict cancer in its early stage.

CONCLUSION

This work proposed a web interface that will collect users input images and diagnose early stage of cervical cancer by using machine learning classifiers.

REFERENCES

Bankar, A., Padamwar, K., & Jahagirdar, A. (2020). Symptom analysis using a machine learning approach for early stage lung cancer. In *2020 3rd international conference on intelligent sustainable systems (iciss)* (p. 246-250). doi: 10.1109/ICISS49785.2020.9315904

Burchell, A. N., Winer, R. L., de Sanjosé, S., & Franco, E. L. (2006). Chapter 6: Epidemiology and transmission dynamics of genital hpv infection. *Vaccine*, *24*, S52-S61. Retrieved from https://www.sciencedirect.com/science/article/pii/S0264410X06005767 (HPV Vaccines and Screening in the Prevention of Cervical Cancer) doi: https://doi.org/10.1016/j.vaccine.2006.05.031

Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., … Mahnashi, M. H. (2021). Skin cancer detection: A review using deep learning techniques. *International Journal of Environmental Research and Public Health*, *18*(10). Retrieved from https://www.mdpi.com/1660-4601/18/10/5479

Hanahan, D., & Weinberg, R. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646-674. Retrieved from https://www.sciencedirect.com/science/article/pii/S0092867411001279 doi: https://doi.org/10.1016/j.cell.2011.02.013

Moldovan, D. (2020). Cervical cancer diagnosis using a chicken swarm optimization based machine learning method. In *2020 international conference on e-health and bioengineering (ehb)* (p. 1-4). doi: 10.1109/EHB50910.2020.9280215

Mridha, K. (2021). Early prediction of breast cancer by using artificial neural network and machine learning techniques. In *2021 10th ieee international conference on communication systems and network technologies (csnt)* (p. 582-587). doi: 10.1109/CSNT51715.2021.9509658

Olusola, P., Banerjee, H. N., Philley, J. V., & Dasgupta, S. (2019). Human papilloma virus-associated cervical cancer and health disparities. *Cells*, *8*(6). Retrieved from https://www

.mdpi.com/2073-4409/8/6/622

Prerita, Sindhwani, N., Rana, A., & Chaudhary, A. (2021). Breast cancer detection using machine learning algorithms. In *2021 9th international conference on reliability, infocom technologies and optimization (trends and future directions) (icrito)* (p. 1-5). doi: 10.1109/ICRITO51393.2021.9596295

Sami, H., Sagheer, M., Riaz, K., Mehmood, M. Q., & Zubair, M. (2021). Machine learning-based approaches for breast cancer detection in microwave imaging. In *2021 ieee usnc-ursi radio science meeting (joint with ap-s symposium)* (p. 72-73). doi: 10.23919/USNC-URSI51813.2021.9703518

Saxena, S., & Prasad, S. N. (2019). Machine learning based sensitivity analysis for the applications in the prediction and detection of cancer disease. In *2019 ieee international conference on distributed computing, vlsi, electrical circuits and robotics (discover)* (p. 1-4). doi: 10.1109/DISCOVER47552.2019.9008083

Sruthi, G., Ram, C. L., Sai, M. K., Singh, B. P., Majhotra, N., & Sharma, N. (2022). Cancer prediction using machine learning. In *2022 2nd international conference on innovative practices in technology and management (iciptm)* (Vol. 2, p. 217-221). doi: 10.1109/ICIPTM54933.2022.9754059