e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 10, October 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 10, October 2024| DOI: 10.15680/IJMRSET.2024.0710073

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

An Empirical Analysis of an AI-Based Cancer Diagnosis Application

Prof. Praveen Kumar K C, Paramashivu, Pooja O S, Niriksha M

Assistant Professor, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

UG Student, Department of ISE, CIT, Gubbi, Tumkur, Karnataka, India

ABSTRACT: One of the most serious illnesses that can affect a person is cancer. The delayed diagnosis may result in worldwide mortality. In order to improve patient care, a wealth of evidence now recommends early detection and treatment initiation for the majority of cancer forms. In this paper, we have considered four important cancers: lung cancer, liver cancer, breast cancer, and cervical cancer. Based on all the datasets collected for the prediction of the types of malignancies we've considered, it compares a number of contemporary algorithms, such as Linear Regression, Logistic Regression, Support Vector Machine (SVM), K-nearest neighbours (KNN), Naive Bayes, Gradient Boosting, Xtreme Gradient Boosting (XGBoost), and Random Forest. Logistic regression had the highest accuracy of 99.6% on the breast cancer dataset, K-nearest neighbour had the highest accuracy of 97% on the lung cancer dataset, Random Forest had the highest accuracy of 99.2% on the cervical cancer dataset, and Random Forest had the highest accuracy of 73.73% on the liver cancer dataset. This will support future research because, as far as we are aware, few studies have compared the most popular algorithms for cancer prediction models. One of the main causes of cancer-related deaths globally is liver cancer, and reducing morbidity and mortality requires early detection and treatment. Although finding and using efficient biomarkers is still a significant problem, biomarkers have the potential to help in the early detection and treatment of liver cancer. Artificial intelligence has become a very promising tool in the field of cancer in recent years, and new research indicates that it holds great promise for enabling the use of biomarkers in liver cancer.

KEYWORDS: cancer, prediction, classification, machine learning (ML), comparison.

I. INTRODUCTION

Cancer is a large group of diseases that can affect any region of the body. It is now the top cause of death worldwide. It develops as a result of a multi-stage process that turns healthy cells into tumor cells. Some early cancers may additionally have signs and symptoms that may be observed, however that is not continually the case. Risk factor means your chances of having the disorder. There is not any way to absolutely ward off cancer but there are ways to lower your chances. Detecting most cancers early often lets in for a better probability of successful remedy. Since the ML model should be useful in predicting cancer, the medical industry stands to gain a great deal from it. In this experiment, we compared eight algorithms on four datasets of distinct tumors in addition to one dataset. In this investigation, we looked at four primary types of cancer: lung, liver, breast, and cervical cancer. Liver cancer is one of the most deadly and quickly spreading cancer kinds. When anything alters the deoxyribonucleic acid (DNA) of healthy liver cells, liver cancer results. Regretfully, you can have very early-stage liver cancer that shows no symptoms. You can probably identify liver cancer in its early stages if you are at risk for it. One type of cancer that arises in the lungs is lung cancer. Globally, lung cancer is the leading cause of cancer-related mortality. Lung cancer is most likely to strike smokers, yet it can strike anyone, even if they have never smoked. Lung cancer usually shows no symptoms or warning indicators in its early stages. However, these don't show up until the sickness is nearing its end. A type of cancer known as breast cancer develops when certain breast cells start to grow abnormally and more quickly than the healthy ones, giving rise to a lump or tumor. Research has shown that a higher risk of breast cancer is associated with environmental, nutritional, and hormonal factors. Breast cancer is most likely caused by a complicated interplay between your genetic makeup and environment.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. LITERATURE SURVEY

Breast cancer is the most common cancer in women globally, and it is growing more common in developing countries where more cases are discovered after the disease has advanced. suggests that prompt screening and treatment of the majority of cancer forms is now necessary, and that early identification helps with patients' clinical therapy. Whether they were in the high-risk or low-risk groups, the group of people with the majority of malignancies was the subject of numerous research that used machine learning and deep learning techniques. According to Austria et al. machine learning techniques are becoming more and more important due to their exceptional capacity to support physicians in making decisions and determining prognoses. The results of this study indicate that there may be several pairs of variables that increase the risk of breast cancer in women. According to Tuncal et al., who employed Support Vector Regression to forecast the incidence of lung cancer in European nations, big datasets lower prediction errors. Gupta et al. made remarkably accurate predictions about the course of lung cancer by combining image analysis with algorithms like KNN and Random Forest. With an emphasis on risk factors like HPV infections, Nithya and Ilango examine how improved feature selection and classification techniques raise the accuracy of cervical cancer predictions. Mudawi and Jaswinder Singh's other research delves into predictive models created especially to tackle the diagnostic difficulties associated with cervical cancer. artificial intelligence (AI) and machine learning (ML) in cancer diagnosis and prognosis. It discusses numerous research that have improved the early identification and treatment of malignancies, including breast, cervical, lung, and liver cancers, by utilizing algorithms like Decision Trees, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and other machine learning techniques. Numerous earlier studies have been conducted in the area of cancer prediction. Numerous machine learning methods, such as decision trees, random forests, SVM, KNN, and others, are employed in prediction. Cruz et al. examined and contrasted the effectiveness of many machine learning algorithms for the prognosis and prediction of cancer. They came to the conclusion that machine learning algorithms will be widely used in healthcare if study quality keeps improving.

III. METHODOLOGY

DATA DESCRIPTION:
In order for someone to utilize the data in the future and comprehend its structure and content, a data description document serves the function of documenting all relevant information about a data file and its contents. The Lung Cancer dataset has 25 columns, 1001 rows, and output levels. The outcomes are "low, medium, and high." 382 times is the low, 356 times is the medium, and 263 times is the high. The output of the liver cancer dataset is "Dataset" and it comprises 584 rows and 11 columns. "1 and 2" are the two results. 2 appears 228 times, while 1 appears 356 times. "Biopsy" is an output of the 859 rows and 36 columns that make up the cervical cancer dataset. "0 and 1" are the two outcomes. There are 824 instances of 0 and 35 instances of 1. The "Diagnosis" result from the 570 rows and 32 columns of the breast cancer dataset. "M and B" represent the two possible outcomes, which stand for benign (not dangerous) and malignant (dangerous). B occurs 357 times, while M occurs 212 times.

DATA PRE-PROCESSING:
A data mining technique called data pre-processing converts unstructured data into a format that is effective and useable. Since data may contain a number of extraneous or missing components, data cleansing is the best course of action.

ALGORITHM:
a)Divide the dataset into training and testing after importing it.
b) Use the training data to train the model;
c) Estimate the value of the testing data;
d) Apply the accuracy score to the testing and forecast data.
To ensure that there would be no issues when the algorithms were applied to the datasets, all of the datasets were first examined and cleansed. Each dataset was then subjected to exploratory data analysis (EDA) to look for trends and conduct a thorough analysis of the data prior to applying the models. Following all of that, we used all eight models on all four datasets to assess the prediction accuracy of each approach. Accuracy was the evaluation parameter that we used to do this.

© 2024 IJMRSET | Volume 7, Issue 10, October 2024| DOI: 10.15680/IJMRSET.2024.0710073

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

EVALUATION PARAMETER:

The evaluation criterion taken into account in this project was accuracy. We determined accuracy by dividing the data into training and testing data, fitting the training data to the model, using a model to predict the test data, then comparing the test result with the expected result to see how many points were correctly predicted.

ACCURACY = TP+TN TP+FP+TN+FN

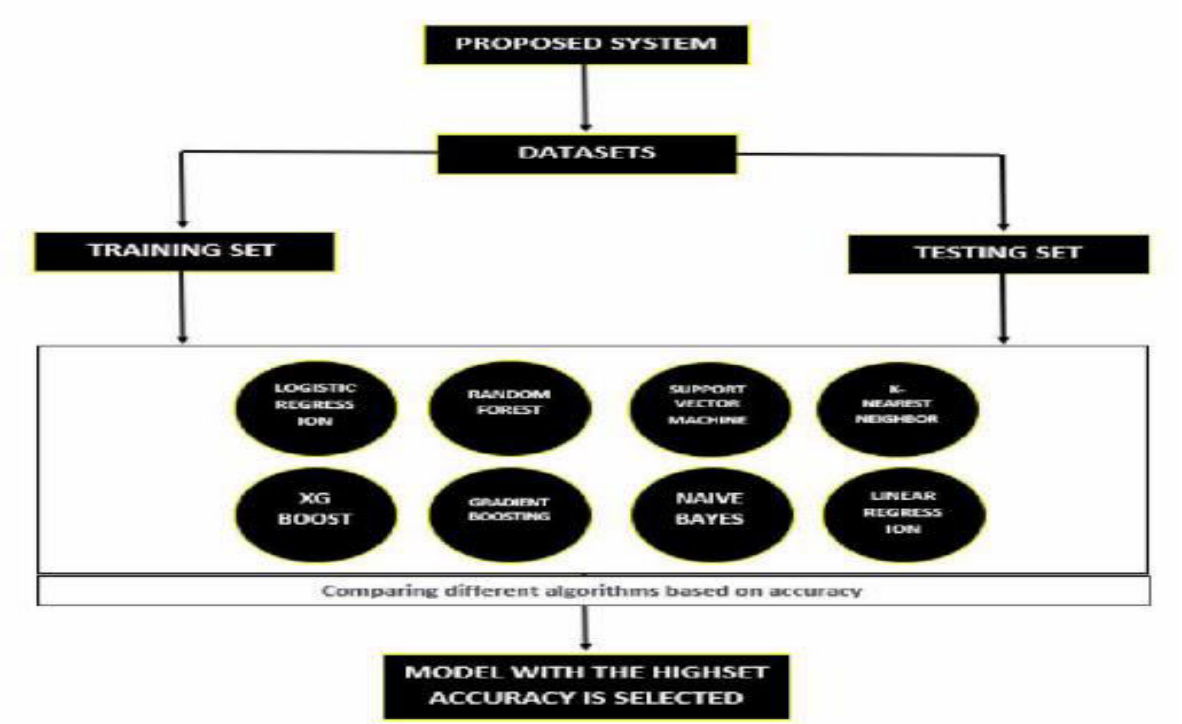

Fig. 1 Working of the Model

IV. RESULTS

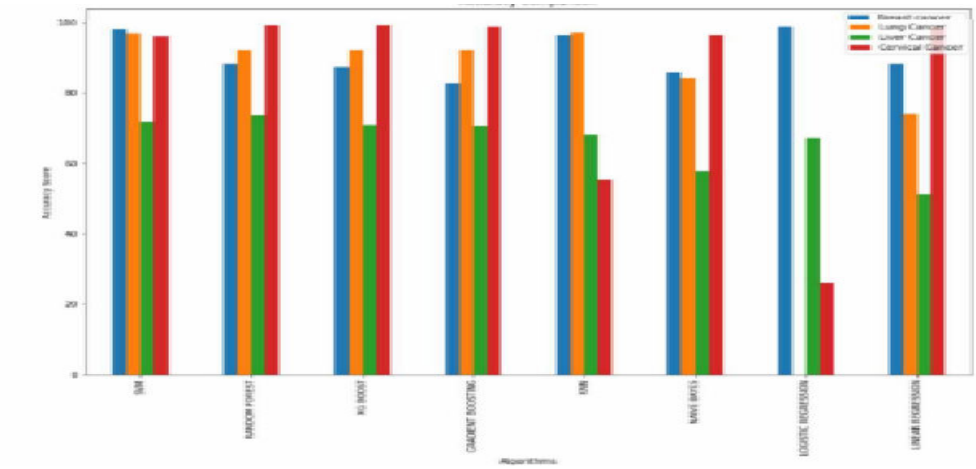

Fig 2. Comparison Chart of 8 Algorithms on all 4 Datasets

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

We think the models we selected will enable medical professionals to make more accurate cancer predictions than they could with conventional techniques. The Lung Cancer model forecasts a patient's excessive, medium, and low risk factors for lung cancer. Within a specific age range of men and women, the version focuses on predicting whether or not the affected individual has liver cancer. By classifying tumors as benign or malignant (cancerous) (noncancerous), the breast cancer model tackles the major barrier to its identification. The version is especially good at predicting the results of biopsy testing to determine whether or not a patient has cervical cancer. According to research, the likelihood of survival increases with the detection of cancer We have attempted to present a comparison study based on the overall performance of several machine learning techniques used in cancer prediction in this review paper. Since machine learning models are getting better than doctors at predicting the onset of cancer, we can try developing a real-life prediction system in the future utilizing the most effective algorithm we have researched. We will also continue to research on all the datasets. surpassing physicians in their ability to forecast when cancer will manifest.. The study evaluates machine learning models for diagnosing four different cancer types: cervical, breast, liver, and lung cancer. The best-performing algorithm for each type of cancer was identified by analyzing the accuracy of each model; Random Forest performed well for liver and cervical cancers, K-Nearest Neighbors for lung cancer, and Logistic Regression for breast cancer. According to the results, customized models can help with early diagnosis and enhance patient outcomes, particularly by detecting cancer early on when therapy is more successful. The study promotes the creation of real-world prediction algorithms to improve clinical diagnostic efficiency and accuracy.

REFERENCES

[1] J. A. Cruz and D. S. Wishart, “Applications of machine learning in cancer prediction and prognosis,” Cancer Inform., vol. 2, 59 77, 2007 [doi: 10.1177/117693510600200030].
[2] A. B. Nassif et al., “Breast cancer detection using artificial intelligence techniques: A systematic literature review,” Artif. Intell. Med., vol. 127, p. 102276, 2022 [doi: 10.1016/j. artmed.2022.102276].
[3] F. J. Shaikh and D. S. Rao, “Prediction of cancer disease using machine learning approach,” Mater. Today Proc., vol. 50, pp. 40 47, 2022 [doi:10.1016/j .matpr.2021.03.625].
[4] N. Rane et al., “Breast cancer classification and prediction using machine learning,” Int. J. Eng. Res. Technol., vol. V9, no. 2, pp. 576-580, 2020. doi:10.17577/IJERTV9IS020280.
[5] Y. D. Austria et al.“Comparison of Machine Learning Algorithms in Breast Cancer Prediction Using the Coimbra Dataset”, International Journal of Simulation: Systems, Science & Technology, vol. 7, no. 10, pp. 23-21, 2019. doi:10.5013/IJSSST.a.20.S2.23.
[6] M. Tiwari et al., 2020, “Breast cancer prediction using Deep Learning and Machine Learning Techniques, SSRN”. Available at:https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3558786
[7] B. Nithya and V. Ilango, “Evaluation of machine learning based optimized feature selection approaches and classification methods for cervical cancer prediction,” SN Appl. Sci., vol. 1, no. 6, pp. 1-16, 2019 [doi :10.1007/s42452-019-0645-7].
[8] N. Al Mudawi and A. Alazeb, “A model for predicting cervical cancer using machine learning algorithms,” Sensors (Basel), vol. 22, no. 11, p. 4132, 2022 [doi:10.3390/s22114132].
[9] J. Singh and S. Sharma, “Prediction of cervical cancer using machine learning techniques,” Int. J. Appl. Eng. Res., vol. 14, no. 11, pp. 2570-2577, 2019. doi:10.3390/s22114132.
[10] K. Tuncal et al., “Lung cancer incidence prediction using machine learning algorithms,” J. Adv. Inf. Technol., vol. 11, no. 2, 91-96, 2020 [doi:10.12720/jait.11.2.91-96].
[11] A. Gupta et al., A Study on Prediction of Lung Cancer Using Machine Learning Algorithms, 2022. doi:10.21203/rs.3.rs1912967/v1.
[12] M. B. S. Ismail, “Lung cancer detection and classification using machine learning algorithm,” Turk. J. Comput. Math. Educ. (TURCOMAT), vol. 12, no. 13, pp. 7048-7054, 2021.
[13] L. Lei et al., “A comparative study of machine learning algorithms for predicting acute kidney injury after liver cancer resection,” PeerJ, vol. 8, p. e8583, 2020 [doi:10.7717/peerj.8583].
[14] N. Tanwar and K. F. Rahman, “Machine Learning in liver disease diagnosis: Current progress and future opportunities,”. In IOP Conference Series, IOP Conf. Ser.: Mater. Sci. Eng. (Vol. 1022, No. 1, p. 012029). IOP Publishing, vol. 1022, no. 1, 2021 [doi:10.1088/1757-899X/1022/1/012029].

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[15] S. Bacha and O. Taouali, "A novel machine learning approach for breast cancer diagnosis," Measurement, vol. 187, p. 110233, 2022 [doi: 10.1016/j .measurement.2021.110233].
[16] M. A. Elsadig et al., "Breast cancer detection using machine learning approaches: A comparative study," Int. J. Electr. Comput. Eng. (2088-8708), vol. 13, no. 1, 2023 [doi:10.11591/ijece.v13i 1.pp736-745].
[17] H. K. Bhuyan and V. Ravi, "'Brahma, B.', & kamila," Health Technol., vol. 12, no. 5, no. K, pp. 987-1005, 2022. Disease analysis using machine learning approaches in healthcare system.doi: 10.1007/s12652-021-03612-z 70
[18] P. J. Ballester et al., "Artificial intelligence for drug response prediction in disease models," Brief. Bioinform., vol. 23, no. 1, p. bbab450, 2022 [doi:10.1093/bib/bbab450].
[19] L. Wu et al., "Machine learning methods, databases and tools for drug combination prediction," Brief. Bioinform., vol. 23, no. 1, p. bbab355, 2022 [doi:10.1093/bib/bbab355].
[20] A. Dang, Cancer Prediction Using Machine Learning Algorithms.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com